

A Network-based Analysis of Polyanion-binding Proteins Utilizing Human Protein Arrays^{*[5]}

Received for publication, November 28, 2006, and in revised form, January 31, 2007. Published, JBC Papers in Press, February 2, 2007, DOI 10.1074/jbc.M610957200

Nazila Salamat-Miller[‡], Jianwen Fang[§], Christopher W. Seidel[¶], Yassen Assenov^{||}, Mario Albrecht^{||}, and C. Russell Middaugh^{‡1}

From the [‡]Department of Pharmaceutical Chemistry and [§]Bioinformatics Core Facility, University of Kansas, Lawrence, Kansas 66047, [¶]Stowers Institute for Medical Research, Kansas City, Missouri 64110, and ^{||}Max Planck Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany

The existence of interactions between many cellular proteins and various polyanionic surfaces within a cell is now well established. The functional role of such interactions, however, remains to be clearly defined. The existence of protein arrays, with a large selection of different kinds of proteins, provides a way to better address a number of aspects of this question. We have therefore investigated the interaction between five cellular polyanions (actin, tubulin, heparin, heparan sulfate, and DNA) and ~5,000 human proteins using protein microarrays in an attempt to better understand the functional nature of such interaction(s). We demonstrate that a large number of polyanion-binding proteins exist that contain multiple positively charged regions, are often disordered, are involved in phosphorylation processes, and appear to play a role in protein-protein interaction networks. Considering the crowded nature of cellular interiors, we propose that polyanion-binding proteins interact with a wide variety of polyanionic surfaces in cells in a functionally significant manner.

Our oldest view of cells as lipid bags filled with proteins and nucleic acids long ago gave way to a much more complex picture in which interior membranes and fibrous proteins create a large variety of unique environments that are critical for individual cellular functions. This picture continues to be refined, but the idea that lipid membranes are the most important structural components defining local environments continues. It is now established that many cellular proteins are capable of interacting with a wide variety of polyanions. We have recently suggested that both the interior and exterior polyanionic surfaces may also have an important role to play in the organization of cellular function. If one views the typically very crowded interior of a cell filled with microtubules, microfilament, ribosomes, lipid membranes, and other structures (1) in the light of the polyanionic nature of these molecular entities (2), the inte-

rior of cells can be seen as a highly polyanionic environment offering an extensive potential interaction surface. Most of our knowledge about polyanion-protein interactions stems from *in vitro* studies that have traditionally been conducted at low concentrations. Under these conditions, the effects of macromolecular crowding and the presence of polyanionic surfaces are neglected. The concept of macromolecular crowding refers to the presence of cell organelles, proteins, polysaccharides, cytoskeleton proteins, and membranes that produces severe steric confinement and a resultant high thermodynamic activity of the molecules present (3, 4). When considering the large surface area and the importance of cellular polyanions such as actin, tubulin, and ribosomes in a crowded cellular environment, the nonspecific electrostatic interaction of proteins containing positively charged regions with these polyanions can be expected. In such cellular matrices, biochemical reaction rates, equilibria, and the diffusion of macromolecules as well as small compounds may be strongly influenced by other macromolecules that do not necessarily take part in more usual stereoscopically determined functional interactions (5–7).

A large number of proteins are commonly recognized as “heparin-binding proteins” (8). Previous work has suggested that the interaction of many of these proteins with polyanions is remarkably nonspecific while maintaining high affinity. The possibility that cellular macromolecules may exert a potential stabilizing influence through nonspecific interactions with proteins *in vivo* has originated from *in vitro* observations that have shown stabilizing effects of polyanions on some proteins (9, 10). These observations have suggested the idea that the extensive array of polyanions in cells might also play some type of organizing role. Functional consequences of such interactions could include the following: (a) bringing proteins together to participate in protein-protein interactions; (b) mediation of protein transport; (c) roles in protein folding as molecular chaperones and stabilizers; and (d) involvement in protein solubility-type diseases. In some cases it is well established that the interactions of proteins with polyanions involve regulatory functions (e.g. binding of growth factors to HS² proteoglycans) (11, 12), genetic information transfer (DNA and RNA), cytoskeleton organization (actin and tubulin microfilaments), chaperone function (9, 13, 14), protein stabilization (10, 15, 16), and non-

* This work was supported in part by Kansas IDeA Network of Biomedical Research Excellence Bioinformatics Core and National Institutes of Health Grant P20 RR016475. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. 1–4, Tables 1–5, and additional references.

¹ To whom correspondence should be addressed: Dept. of Pharmaceutical Chemistry, University of Kansas, 2030 Becker Dr., Lawrence, KS 66047. Tel.: 875-864-5813; Fax: 875-864-5814; E-mail: middaugh@ku.edu.

² The abbreviations used are: HS, heparan sulfate; PABP, polyanion-binding protein; MES, 2-(N-morpholino)ethanesulfonic acid; PCE, protein continuum electrostatics; ORF, open reading frame.

Polyanion-binding Protein Networks in the Human Proteome

classical transport of proteins within and without cells (17–19). As mentioned above, it also appears that cellular polyanions may play a critical role in protein aggregation diseases (20–23).

To test this general hypothesis, we previously examined the extent of intracellular polyanion-protein interactions. Using COS-7 cells and a combination of polyanion pulldown experiments and two-dimensional gels as well as antibody arrays, we found that hundreds to thousands of such interactions appear to take place (2). Many of these PABPs were identified, and recognition that a substantial number of these proteins are “natively unfolded” was noted, and the nature of typical polyanion-binding sites was better defined.

In a much more detailed study, we used yeast protein arrays to identify the interaction of the polyanions actin, tubulin, heparin, HS, and DNA with thousands of proteins. An enrichment of several hundred PABPs that were unstructured as well as those possessing kinase substrate activity was observed (24). The charged nature of the polyanion-binding sites was further clarified, and a network analysis was performed. It was found that many PABPs appear to participate in a variety of cellular protein-protein interaction networks.

In this study, we perform a similar analysis employing a human protein array, and we provide further evidence for the existence of an extensive role for polyanion-protein interactions in the activities of much more complex eukaryotic cells.

EXPERIMENTAL PROCEDURES

Lyophilized G-actin from bovine muscle (43 kDa), heparin, HS from porcine intestinal mucosa with a mean molecular mass of 18 and 14 kDa, the metachromatic dye Azure A, and calf thymus double-stranded DNA were purchased from Sigma. Bovine brain tubulin (100 kDa) was a kind donation from Dr. Richard Himes, Division of Biological Sciences, University of Kansas. EZ-link biotin-LC-hydrazide, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, MES, and EZ-link psoralen-PEO₃-biotin were obtained from Pierce. Human protein array kits, including buffers and controls, were purchased from Invitrogen. All protein arrays were scanned using a Molecular Devices GenePix 4000B (Sunnyvale, CA) at 635 nm with 10- μ m pixel size.

Biotinylation of Polyanions

Actin and Tubulin—Actin was dialyzed against 10 mM phosphate-buffered saline overnight to avoid a cross-reaction between the Tris salt contained in its lyophilized form and the biotinylation reagent. Tubulin was dialyzed against two exchanges of 10 mM phosphate-buffered saline for 3 h at 4 °C. Actin and tubulin (~2.0 mg/ml) were biotinylated according to the manufacturer's instructions using the biotin-XX-sulfosuccinimidyl ester (5 nmol/ μ l) included in the Invitrogen protein array kit. Excess biotin was removed using a gel filtration resin provided in the kit. The biotinylation efficiency was assessed by performing Tris-glycine SDS-PAGE and a Western blot of the biotinylated polyanions and the provided reference proteins according to standard protocols. Detection and visualization of samples were performed using a streptavidin-alkaline phosphatase conjugate and a chemiluminescent substrate, respectively.

Heparin and HS—Compounds were dissolved in 0.1 M MES buffer to a final concentration of ~4.0 mg/ml. The biotinylation reaction was conducted by addition of 25 μ l of biotin hydrazide in dry Me₂SO and 12.5 μ l of freshly prepared 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, with final concentrations of 50 and 5.0 mM, respectively, to the MES/heparin or MES/HS solution. Solutions were stirred overnight at room temperature. The carbodiimide chemistry coupled biotin-LC-hydrazide to heparin and HS through uronic acids (carboxylate groups) (25). Excess biotin was removed using either dialysis cassettes (molecular weight cut-off of 3,500) or polyacrylamide desalting columns (Pierce). Biotinylation efficiency was assessed by performing a dot blot on nitrocellulose membranes (Schleicher & Schuell). Approximately 2 μ l of each biotinylated sample was placed on the membrane, and the blot was air-dried. Biotinylated tubulin or bovine serum albumin (both at a molar ratio of 9:1) and nonbiotinylated polysaccharide or buffer were used as positive and negative controls, respectively. This procedure was repeated three times, followed by blocking and washing of the membrane according to the Western blot protocols. Detection and visualization of samples were performed as above.

Calf Thymus DNA—A final concentration of 200 μ M EZ-link psoralen-PEO₃-biotin in doubly deionized H₂O was added to a 1 mg/ml solution of DNA in TE buffer (10 mM Tris, 1 mM EDTA, pH 7.4) after boiling and immediate cooling of the DNA solution. The reaction centrifuge tube was placed on ice and irradiated with a long wavelength UV source (Mini Ralight lamp, UVP Inc., San Gabriel, CA) for 40 min. The biotinylated solutions of DNA were stored at –20 °C. Detection and visualization of biotinylated DNA were performed through a dot-blot procedure as described above.

Quantification of Polyanions

Actin and Tubulin—Final concentrations were determined by UV absorbance spectroscopy at 290 nm for actin and 280 nm for tubulin, using extinction coefficients of 2.66×10^4 (26) and 1.15×10^5 M^{–1}cm^{–1} for actin and tubulin, respectively, and a cell path length of 1.0 cm.³

Heparin and HS—For concentration determination of carbohydrates, azure A, a metachromatic dye, was utilized (27). A stock solution of Azure A was prepared in 0.2% (v/v) aqueous formic acid, pH 3.5, at a final concentration of 5.04×10^{-5} M. One ml of the dye stock solution was titrated with 1–10 μ l of 1.0 mg/ml carbohydrate solution prepared in phosphate-buffered saline and vortexed (28). The absorbance of each sample was measured at 620 nm, where the extent of disappearance of Azure A was followed as a function of carbohydrate concentration. In each case, the nonbiotinylated heparin and HS were used as standards for the biotinylated samples. The absorbance spectroscopy confirmed that Azure A was bound to both biotinylated and nonbiotinylated heparin and HS. These experiments were conducted after observing that biotinylated HS failed to bind to another metachromatic dye, 1,9-dimethylmethylene blue. This is in contrast to the nonbiotinylated form that shows binding to this dye.

³ T. Mitchison, personal website.

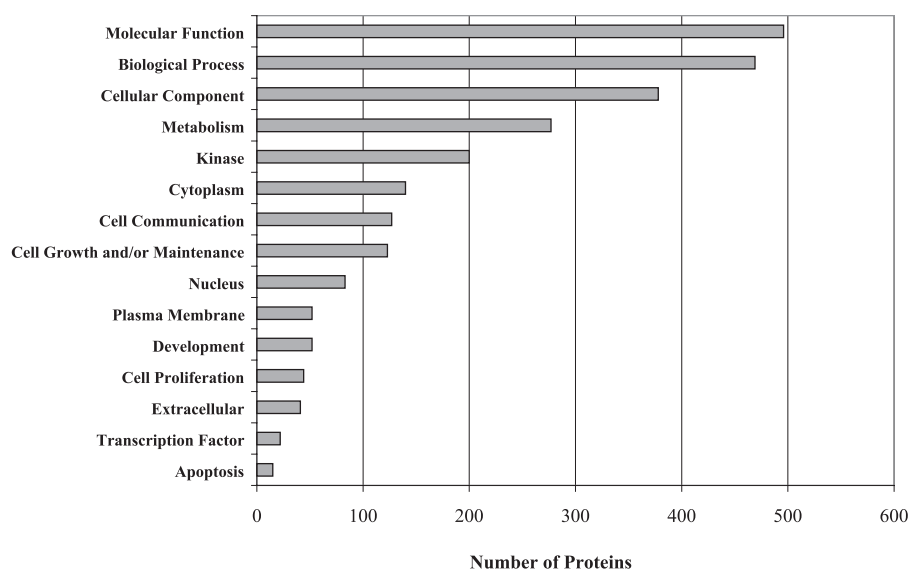


FIGURE 1. Gene ontology classification of the human proteins printed on the array (copyright 2005, Invitrogen, used with permission).

Calf Thymus DNA—DNA was quantified using an optical density of 1.0 for a 50 $\mu\text{g/ml}$ solution at 260 nm.

Human Protein Array Experiments

The human arrays employed contain 4,985 doubly spotted proteins on a proprietary polymer coating. These proteins were derived from the human UltimateTM ORF Clone Collection (Invitrogen) and expressed in insect cells using a baculovirus expression system as N-terminal glutathione *S*-transferase fusion proteins (29). The human proteins are classified by the Gene Ontology consortium and contained a range of proteins from the cytoplasm, nucleus, plasma membrane, and extracellular matrix. These proteins are involved in variety of functions, including cell communications, cell growth and/or maintenance, cell proliferation, metabolism, apoptosis, development, and transcription (Fig. 1).

Experiments were conducted according to the manufacturer's protocol. Briefly, protein arrays were probed with 120 μl of 50 $\mu\text{g/ml}$ of each polyanion in the presence and absence of 0.15 M NaCl in the probing solution at pH 7.4. After subsequent washings, the arrays were exposed to streptavidin-Alexa Fluor[®] 647. The arrays were then washed, dried, and scanned with the GenePix[®] 4000B at 635 nm to acquire a fluorescence image. The resultant files were produced using GenePix[®] Pro 5.0 software. Each image was analyzed using the appropriate grid downloaded from Invitrogen. For all of the scanned images, a program available through Invitrogen (ProtoArrayTM Prospector, version 3.1.1) was used to obtain the identity of the proteins that demonstrated significant interaction with the probes.

Interactions were considered statistically significant when they displayed a mean signal greater than the median signal of all protein spots on the array plus one times the standard deviation (+1S.D.). This +1S.D. threshold has a Z-score that is greater than 1. For such threshold values, polyanion-binding proteins with low, medium, and high confidence levels are included in the data set. Proteins that bound the polyanions at the +2 and +3S.D. level demonstrate a higher confidence level.

However, we decided to generally include the greatest number of proteins involved in any type of binding to polyanions (specific or nonspecific), and we therefore chose to study PABPs at the level of +1S.D., unless otherwise indicated.

Searching for Enriched Amino Acids in the Sequence of PABPs Using MATLAB (The Math Works, Inc., Version 7.0.4)

An in-house MATLAB program was used to search for basic amino acids (lysine, arginine, and histidine) in human PABPs and to calculate their frequency in percent for each protein. Additionally, this program was used to analyze the percent of basic amino acids in the overall population of human proteins on the array.

The sequence of each protein was supplied by Invitrogen. We employed a one sample *t* test to statistically analyze the results between the means of one particular amino acid percentage for proteins that interacted with the polyanionic probes (PABPs) and the same amino acid in a general population set. Similarly sized 1,000 random sets of general population proteins were generated, and the percent of basic amino acids was also calculated for these sets. In addition, we investigated whether the enrichment of positive residues was more pronounced within a sequence locally. To this end, we used a perl script to calculate the number of positive pitches in human PABPs as well as the entire protein population on the array. Positive pitches were defined as a sequence run containing three or more positive residues out of any five continuous residues.

Isoelectric Point Calculations

The program code for theoretical pI calculations was obtained through personal communications with the code author.⁴ This program was used to generate and compare the theoretical pI values for the identified set of PABPs as well as the human protein population on the array. The results of the PABPs and general population sets were compared utilizing a one-sample *t* test.

Macroscopic Electrostatic Models Using Protein Continuum Electrostatics

By employing an in-house MATLAB program, the list of human PABPs generated from the human array experiments was compared with a list of human proteins with solved three-dimensional structures in ExPASy server. Consequently, a PCE web tool based on macroscopic electrostatics with atomic details was used to calculate the electrostatic surface potential of these structures. The PCE server calculates surface potentials and electrostatic energies via an infinite differences solution to

⁴ C. Putnam, personal communication.

Polyanion-binding Protein Networks in the Human Proteome

the Poisson-Boltzmann equation (30). Default values of 4.0, 80.0, and 0.1 were used for protein internal and solvent dielectric constants and ionic strength, respectively.

Searching for Unique Motifs (Sequence Signatures) in PABPs using MEME and InterPro

Sequence Signature Analysis—A sequence signature is defined as a highly conserved region, a recurring theme or pattern that is found in a group of related sequences. By this definition, a sequence signature could be a protein family, functional domain, functional site, or any conserved region of unknown function. Thus, the actual physical manifestation of a signature can vary greatly in size (31). In this study, signature analysis was performed for the sequences of the human PABPs binding to each polyanion probe using MEME (version 3.0.10) (32). MEME implements an unsupervised learning algorithm for discovering gapless signatures in a group of related protein sequences. We chose the maximum number of signatures to be 50 and the minimum and maximum width of the signatures to range from 5 to 300 amino acids, respectively. The *E* value cut-off was set to 0.1 to report only statistically significant results. In addition, the type of distribution was set to 0 or 1 occurrence per sequence, and the minimum number of sites was set to 3. Therefore, only signatures that were shared by three or more PABPs were discovered by the MEME search.

Sequence Signature Characterization—The consensus sequence of each signature model was searched for in all InterPro member data bases using a local version of InterProScan. InterPro is an integrated collection of the most commonly used data bases of protein families, domains, and functional sites (33). The program InterProScan allows a user to search for sequence signatures in any number of these data bases simultaneously.

Human Kinases and PABPs

The list of known human protein kinases (kinome) was obtained from the supplemental information published by Manning *et al.* (34) and was compared with the list of human PABPs and non-PABPs (control). All of the phosphorylated proteins are represented in official gene symbol format in this data set. We used a BatchSearch tool to transform our accession numbers to gene symbols. Unfortunately, not all proteins that are represented by accession numbers possess known gene symbols. Thus, we were unable to identify gene symbols for all of the human PABPs. The statistical analysis of the results was conducted using Fisher's exact test.

Phosphorylation Sites in PABPs

Phospho.ELM is an experimentally verified data base of the phosphorylation sites in eukaryotic proteins (35). This data set (version 5.0) contains 2,540 substrate proteins from different species covering 7,206 phosphorylation sites at tyrosines, serines, and threonines. A query was made between proteins in this data set and the human PABPs to obtain the number of phosphorylated PABPs. A similar analysis was also performed between this data base and the entire pool of human proteins on the array. All of the phosphorylated proteins are represented in UniProt nomenclature in this data set. Accession number

transformation of PABPs to this appropriate format was performed using a BatchSearch tool. The statistical analysis of the results was conducted as above using Fisher's exact test.

PABPs as Disordered Proteins

DisEMBL was utilized for a sequence analysis of PABPs to estimate the extent of their disorderliness (36). Based on artificial neural networks, which are trained on three different data sets, this program predicts several definitions of disorder. Because no formal definition for protein disorder exists, we decided to use the "hot loops" definition to calculate the percent disorderliness for human PABPs as well as non-PABPs. These loops are defined as sections of a protein with a high degree of mobility as determined from C- α temperature factors (*B*-factors) (36). An unpaired, two-sampled unequal variance *t* test was used to investigate the statistical difference between the means of percent hot loops for PABPs and non-PABPs.

Interaction Networks of PABPs

A human interactome was compiled by combining predicted protein-protein interactions from HiMAP from human physical protein interaction networks in IntAct, the human protein reference data base, BioGrid, and the data base of interacting proteins. Because these data sets use different accession numbers, we have converted all to NCBI Entrez Gene identifiers, leading to slightly different numbers of PABPs in the reported tables. To probe specific topological properties of sub-networks formed by PABP data sets, we investigated the number of protein-protein interactions involving at least one PABP.

We constructed 50,000 random sub-networks by arbitrarily picking the same number of proteins in the respective PABP data set and including their interaction partners. This is represented as a histogram in which the horizontal axis displays the different values obtained for the number of interactions in the generated sub-networks, and the vertical axis shows how many random sub-networks contain the given number of interactions. We used the resulting histogram of the number of interactions contained in each random sub-network to estimate a *p* value quantifying the significance of the sub-network for the respective PABP data set. Additionally, a 95% binomial confidence interval for the *p* value was calculated. To give a concrete example of our analysis, there were 149 PABPs in the data set detected at the statistical level of +3S.D., and 90 of them were present in the compiled human interactome. Thus, we generated 50,000 different random sub-networks of 90 proteins and their direct interaction partners. Then we counted the protein-protein interactions in each sub-network to derive a histogram. In sub-networks formed exclusively by PABPs, we additionally investigated the number of clusters of interacting proteins, also referred to as *connected components* in graph theory (37). Two interacting proteins belong to the same cluster if and only if there exists a path of interactions between the proteins. Similar to the case of the topological property described above, a *p* value was estimated for each PABPs data set by generating 50,000 random sub-networks and computing a histogram for the number of clusters in them. Every *p* value estimation includes the calculation of a 95% binomial confidence interval.

We have also determined the yeast homologs of human PABPs (see below). Thus, for each human PABP data set, we could use a corresponding set of homologous yeast proteins. Again, we have performed the topological analyses described above on the resulting sets of yeast proteins using the compiled yeast interactome discussed in our previous paper (24).

Comparing Human and Yeast PABPs

Identifying Orthologs—We inquired about the PABPs, which are common between human and yeast, and used the Inparanoid data base (38, 39) to accomplish this task. This program detects orthologs based on all-*versus*-all sequence comparison between two species. The stand-alone Inparanoid program (version 1.35) was downloaded. This data base consists of 22,216 and 5,777 analyzed proteins for human and yeast, respectively. The program was used with the default settings with a score cutoff of 50 bits and a sequence overlap cutoff of 0.5. The ortholog sequence pairs with a bootstrap value of 100% and a score of 1.00 in each cluster was considered.

Homolog Search—Inparanoid was used to discover ortholog clusters. BlastP (version 2.2.14), on the other hand, was used to determine how many homologs are present in yeast and human and whether their number is significantly different from random data sets. One hundred random, nonredundant data sets were generated using MATLAB for bootstrap analysis of the homologs. An *E*-value cutoff of 0.1 was utilized to accept or reject the results of Blast.

Domain Search—The presence of significant biological domains was investigated in both yeast and human PABPs using a local server containing the Pfam data base (40). Version 20.0 currently has 8,296 entries.

RESULTS AND DISCUSSION

Biotinylation of Polyanions—The successful biotinylation of the polyanion probes was confirmed through Western and dot blot analysis. The different panels in supplemental Fig. 1 demonstrate the Western blots of biotinylated actin and tubulin (at a molar ratio of 9:1) as well as the biotinylated standards used in these experiments. Dot blots for heparin, HS, and DNA also verified their biotinylation.

Quantification of Heparin and HS—Different concentrations of heparin and HS were quantified by using Azure A. This dye, in the absence of heparin and HS, exhibited an absorbance peak maximum at ~630 nm (supplemental Fig. 2), which shifts to lower wavelength in the presence of these macromolecules. Detailed results of this procedure are reported in the supplemental material and our previous publication (24).

Protein Array Experiments—To identify PABPs in the human proteome, we utilized human protein arrays containing 4,985 human proteins and probed them with the polyanionic biotinylated actin, tubulin, heparin, HS, and DNA. Approximately 800 positive hits were detected with a total of 397 different human proteins identified as PABPs (see Table 1). Twenty four of the PABPs interacted with all of the probes. Additionally, significant overlap was observed in which some PABPs interacted with two or more probes. Specifically, it was found that 69, 62, and 38 proteins interacted simultaneously with two, three, and four probes, respectively. The standard ORF name

TABLE 1

The number of human proteins that interacted with each of the five polyanion probes

Polyanion	PABPs on human array
Actin	202
Tubulin	90
Heparin	194
HS	143
DNA	170

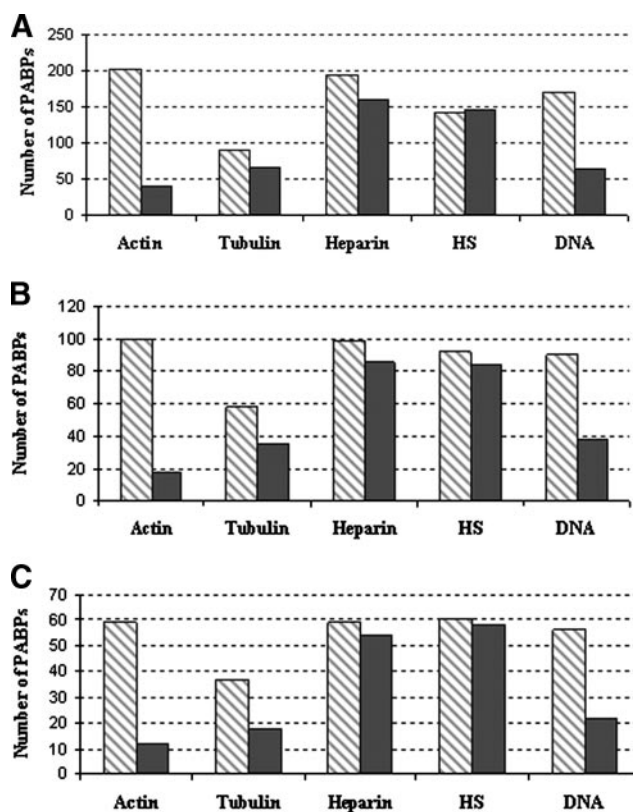


FIGURE 2. Effect of NaCl on the extent and number of PABPs interacting with human proteins on the arrays at the three statistical levels of 15.D. (A), 25.D. (B), and 35.D. (C). Patterned and solid bars represent the number of proteins obtained from the experiments conducted in the absence and the presence of 0.15 M NaCl, respectively.

(GenBank™ accession number) of all proteins that interacted with each polyanion probe is presented in supplemental Table 1. These proteins are deposited in the NCBI gene expression and hybridization array data repository (GEO, www.ncbi.nlm.nih.gov) with series record number of GSE6354.

To evaluate the contribution of electrostatic interactions to binding, we conducted array experiments in the presence of 0.15 M and zero salt (NaCl) concentrations. The results are shown in Fig. 2. As expected, the number of PABPs was significantly reduced when the arrays were exposed to actin, tubulin, and DNA in the presence of NaCl. This effect was not as pronounced in the case of heparin and HS where little difference is seen in the presence and the absence of NaCl. Although electrostatic interactions between PABPs and polyanions are well documented (41, 42), the extent of such interactions has not been extensively characterized. In fact, in previous studies of the yeast proteome, we found even less evidence for electrostatic interaction between PABPs and the same polyanions based on inhibition of binding by even higher salt concentra-

Polyanion-binding Protein Networks in the Human Proteome

TABLE 2

The average percent of basic amino acids generated by MATLAB from the experimental PABPs (test set) and the overall population of proteins on the human array

Polyanion	Mean % Lys	Mean % Arg	Mean % His	Mean % all positive residues	Mean pI
Actin	8.23	7.25	2.65	18.13	8.35
Tubulin	8.29	7.64	2.67	18.60	8.83
Heparin	8.30	7.33	2.48	18.11	8.34
HS	9.05	7.81	2.50	19.36	8.93
DNA	8.17	7.32	2.35	17.84	8.64
Average				18.41	
Population	6.16	5.69	2.50	14.35	7.22
Difference				4.06 (28.3%)	
p value	0.0002 ^a	>0.0001 ^a	0.638	0.0001 ^a	0.0003 ^a

^a Value indicates a significant difference ($p < 0.05$).

tions (24). Note that it may be possible to disturb such interactions with higher concentrations of NaCl, although disruption of the structure of the individual proteins becomes increasingly possible. Nevertheless, this simple experiment does strongly suggest the presence of non-columbic interactions, which may contribute to protein-polyanion interactions. This is also consistent with other observations (41, 42).

Searching for Positively Charged Amino Acids in the Sequence of PABPs Using MATLAB—It is well established that the interactions between PABPs and polyanions is at least partially electrostatic in nature (41, 42) (see above). We therefore computationally investigated this fact at the human proteome level. We searched for Lys, Arg, and His residues in the identified PABPs and in random sets of sequences. The percentages found are shown in Table 2. These values were generated by calculating the fraction of the basic amino acids in each human protein that interacted with polyanions (PABPs) and the overall population of proteins present on the arrays, respectively. These values were subsequently averaged over the entire range of each category of proteins. The results from a one-sample *t* test demonstrated a significantly higher percentage of Lys and Arg but not His in the PABPs. These data are in contrast to the results from the yeast PABPs in which Arg did not demonstrate a significant increase in occurrence in PABPs, and His displayed only a marginal increase (24). Lys residues also displayed a significant increase in frequency in the yeast PABPs. Overall, a 28% increase is observed for positively charged residues in PABPs compared with the general population set of all human proteins on the array.

It is known that many polyanion-binding proteins are studied with positive patches that at least partially constitute the polyanion-binding site (41). Therefore, we utilized a perl script to estimate the number of such positive patches in the PABPs identified here. The results of this investigation demonstrated that the mean number of positive patches for the total population proteins on the array was 2.68. This value was statistically greater ($p = 0.0051$) for the actin (5.23), tubulin (4.54), heparin (5.40), HS (4.89), and DNA (4.64)-binding proteins. Such conserved positive charges on PABPs are often expressed by polybasic domains in the structure of these proteins and are important for their function. For instance, many signaling molecules (e.g. K-Ras and Rac1) are shown to be associated with the lipid membranes through electrostatic interactions of their charged domains (43).

pI Calculations—We demonstrated above that PABPs are enriched in basic amino acids when compared with the general

population set of proteins on the human protein microarrays. Such enrichment would be predicted to be manifested in the pI values of such proteins. The pI values obtained for the five sets of PABPs and the entire human population of proteins, averaged over the total number of each set, exhibited a statistically significant difference ($p < 0.05$; see Table 2).

Macroscopic Electrostatic Models Using PCE—To visually probe the spatial location of the basic amino acids, we employed a PCE approach. A search for the solved crystal structures of human PABPs resulted in 17 proteins found in the Protein Data Bank. Electrostatic maps of human PABPs with solved Protein Data Bank crystal structures demonstrated the consistent presence of well delineated and positively charged surface patches (supplemental Fig. 3, depicted in blue). In some structures, the polyanion-binding sites are more pronounced than in others (for example in Protein Data Bank codes 1MJM and 1I0Z compared with 1T0P). It can be speculated that one protein that contains several positive patches may simultaneously interact with several polyanionic sites *in vivo* for reasons such as improved stability (see PABPs as disordered proteins below) or perhaps functional regulation (see phosphorylation sites in PABPs below). More detailed information about these structures can be found in supplemental Table 2.

Searching for Unique Motifs (Sequence Signatures) in PABPs Using MEME and InterPro—To investigate whether there are any other common features in PABPs in addition to their unique charge distributions, a search for unique sequence signatures was conducted. We identified 26, 9, 23, 14, and 15 specific motifs for actin, tubulin, heparin, HS, and DNA-binding proteins using MEME analysis. These features matched existing entries in the InterPro data base. Most of them belong to expected categories such as RNA and DNA binding domains. Protein kinase and growth factor motifs (e.g. heparin-binding, fibroblast, interleukin 1, etc.) were among the categories in PABPs, which appear to interact with all five probes. The signature sequences identified by MEME and their InterPro analysis are available in the supplemental material and on line.

Human Kinases and PABPs—Manning *et al.* (34) have cataloged the protein kinase complement of the human genome. They have identified 518 genes as putative protein kinases, which is about 1.7% of all human genes (34). Because protein kinases facilitate many signal transduction events in eukaryotic cells, we investigated whether any of the human PABPs can be identified as kinases. One rationale for this investigation stems from the fact that kinases catalyze the attachment of phosphoryl groups through the use of ATP to specific amino acids on

TABLE 3

This table lists the *p* values estimated for the number of human protein-protein interactions involving PABPs and the number of clusters. The *p* values show the significance of the corresponding sub-network for each statistical PABP detection level (defined by the standard deviation).

PABP type	S.D.	No. of PABPs	PABPs present in the interactome	PPI involving PABPs (<i>p</i> value)	Clusters of PABPs (<i>p</i> value)
Actin binding	1	189	141	0.072 ± 0.002 ^a	0.11728 ± 0.0028 ^a
Actin binding	2	96	73	0.397 ± 0.004	0.18240 ± 0.0034
Actin binding	3	58	44	0.228 ± 0.004	1.00000 ± 0.0000
DNA binding	1	161	123	0.033 ± 0.002	0.00008 ± 0.0001
DNA binding	2	85	64	0.038 ± 0.002	0.00094 ± 0.0003
DNA binding	3	54	44	0.181 ± 0.003	0.02082 ± 0.0013
Heparin binding	1	184	124	0.006 ± 0.001	0.00018 ± 0.0001
Heparin binding	2	92	64	0.004 ± 0.001	0.00308 ± 0.0005
Heparin binding	3	54	40	0.003 ± 0.000	0.07490 ± 0.0023
HS binding	1	133	97	0.124 ± 0.003	0.00120 ± 0.0003
HS binding	2	83	62	0.410 ± 0.004	0.00008 ± 0.0001
HS binding	3	54	38	0.64 ± 80.004	0.30993 ± 0.0041
Tubulin binding	1	86	65	0.42 ± 30.004	0.03768 ± 0.0017
Tubulin binding	2	55	41	0.90 ± 50.003	1.00000 ± 0.0000
Tubulin binding	3	36	26	0.735 ± 0.004	1.00000 ± 0.0000
All	1	370	264	0.003 ± 0.001	0.00002 ± 0.00004
All	2	214	156	0.009 ± 0.001	0.00002 ± 0.00004
All	3	149	111	0.016 ± 0.001	0.00080 ± 0.0002

^a S.D. indicates 95% confidence level.

target proteins. Therefore, kinases by their very nature are polyanion (ATP)-binding proteins. A comparison between our experimentally derived list of human PABPs and the Manning *et al.* (34) list of human kinases resulted in a significantly larger number of PABPs that were kinases compared with the population of proteins on the human arrays. Twenty five proteins out of a total of 392 human PABPs were identified as kinases (6.4%). For non-PABPs, 127 proteins out of the total population of human proteins (3.2%) were kinases. Note that we were not able to exactly match all human PABPs and array proteins in the Manning list. Fisher's exact test demonstrated a *p* value of 0.002, indicating that kinases, as a category of polyanion-binding proteins, not only interact with ATP but also showed some indiscriminate behavior toward other polyanions such as the ones examined in this study. This result should also be considered in the context of new findings that terminal kinases (kinases at the end of signaling cascades) are found to physically occupy target negatively charged genes upon activation in yeast (*Saccharomyces cerevisiae*) (44).

Phosphorylation Sites in PABPs—In addition to investigating the relationship between the human kinases and PABPs, we further explored a potential relationship between phosphorylation sites in eukaryotic proteins and human PABPs. This ensured that the entire range of kinase substrates as well as kinases was considered. We hypothesized that if PABPs play a role in the regulation of cellular events or perhaps use polyanions as stabilizing partners, they might be more susceptible to this particular chemical modulation than non-PABPs. Our results demonstrated that 136 out of a total of 333 (41%) of the PABPs contained phosphorylation sites. Not all accession numbers of PABPs possessed equivalent UniProt codes. This was only 23% for non-PABPs (906 non-PABPs proteins were identified from a pool of 3,925 proteins on the array in the UniProt format). By using Fisher's exact test, this number was found to be significantly smaller than the phosphorylation sites in PABPs (*p* = 6.7 × 10⁻¹²).

PABPs as Disordered Proteins—Numerous recent investigations suggest that the relationship between the three-dimensional structure of a protein and its biological function(s) may

need to be re-evaluated (45, 46). This idea originates from new evidence for functional roles of disordered regions in proteins (47, 48). Because polyanions have been shown to stabilize some proteins *in vitro* and promote their folding, we hypothesized that they may serve as potential interaction partners for "disordered" proteins to stabilize them *in vivo* (2, 9). We have focused on the hot loops type of disorder (defined as coils/loops with a high degree of mobility as determined from crystallographic *B*-factors), because this identifier has demonstrated as a more accurate performance evaluation (36). The results of such an analysis for PABPs and non-PABPs demonstrated that 39% of the PABPs contained hot loops. This value was only 27% for the complete human protein population. The *p* value for this comparison from the unpaired, two-sampled unequal variance *t* test was much lower than 0.001, demonstrating a significant increase in the disordered nature of (some) PABPs. It has also been demonstrated that 83–94% of transcription factors (DNA binding and thus polyanion-binding proteins) possess extended regions of disorderliness (48). The question, however, remains as to what extent such disorderliness is possible *in vivo*, where there are a substantial number of polyanionic surfaces in close proximity to most PABPs. We propose that the phrase "natively disordered" may be inappropriate, at least *in vivo*, on the basis of our proteomic investigations as well as studies by others that have shown, for example, that the disordered FlgM protein gains structure in living cells (49).

Interaction Networks of PABPs—To investigate whether human PABPs are involved in some type of protein (polyanion) network as seen in yeast (24), we compiled a human interactome consisting of 97,157 physical interactions between 18,461 proteins. We addressed the following two questions. 1) Do polyanions and PABPs appear to play important functional roles within these networks? 2) Are PABPs highly interconnected, forming some type of network of physical interactions among themselves? If the answer to the first question is positive, then PABPs should be involved in significantly more protein-protein interactions than a random set of proteins. We found that the number of interactions involving all PABPs is indeed significantly larger than in random sub-networks as

Polyanion-binding Protein Networks in the Human Proteome

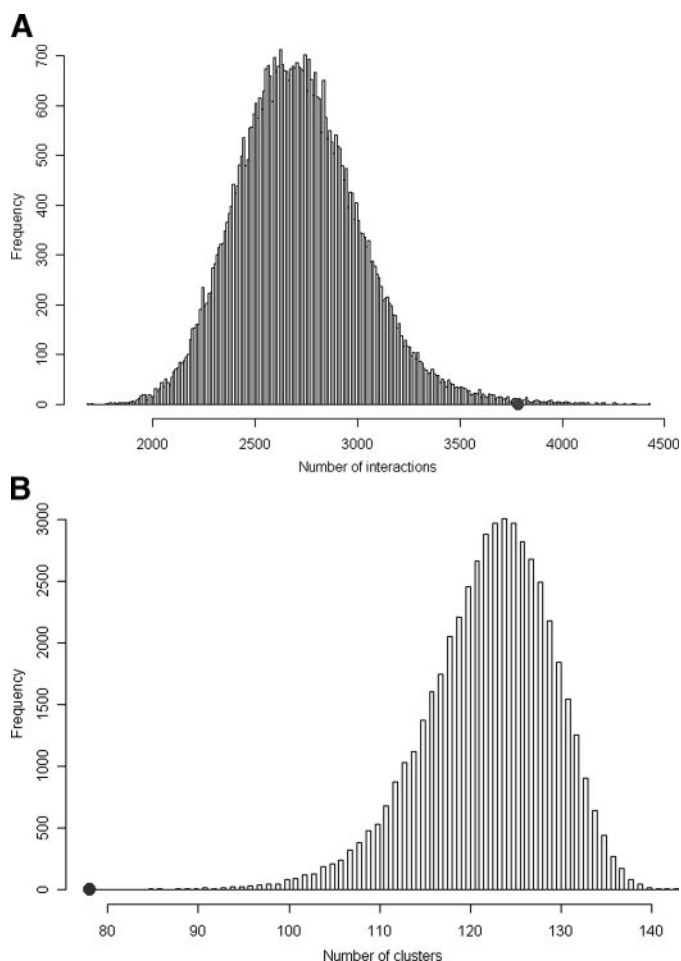


FIGURE 3. In both sections of this figure only the histograms at the level of 1S.D. are shown. A, histograms for the number of human protein-protein interactions in random sub-networks. The horizontal axis of the histogram represents the different values obtained for the number of interactions in the generated sub-networks. The vertical axis shows how many random sub-networks contained the given number of interactions. These histograms are produced for three different statistical levels +1S.D., +2S.D., and +3S.D. of PABP detection to estimate the *p* values shown in Table 3. The black circle represents the number of interactions in the sub-network formed by PABPs and their direct interaction partners. B shows the histograms for the number of clusters in yeast protein-protein interactions involving yeast proteins that are homologous to human PABPs in random yeast sub-networks. These histograms are also produced for three different statistical levels +1S.D., +2S.D., and +3S.D. of PABP detection to estimate the *p* values shown in supplemental Table 3. The black circle represents the number of clusters in the sub-network formed by PABPs only.

judged by the estimated *p* values (Table 3). For instance, at the statistical level of +1S.D., the *p* value from the histogram is 0.003 (Fig. 3A), indicating that PABPs are involved in protein-protein interactions more than by chance. Concerning the second question, a high connectivity between PABPs should be reflected by few large clusters of interacting PABPs; this means that PABPs should form several coherent and connective cluster(s) and only a few disconnected small ones. Indeed, the results of our study show that the network of all PABPs and some sub-networks of PABPs, such as the DNA-binding and heparin-binding species, tend to be connected, usually having larger clusters than randomly generated sub-networks of the same size (Table 3). This can be seen in histogram A of supplemental Fig. 4 where the *p* value is 0.00002 at the level of +1S.D.

The inclusion of predicted interactions into the compiled human network was necessitated by the limited availability of experimental data on human protein-protein interactions, which results in a partially incomplete human interactome. This incompleteness can be observed by the relatively large fraction of 106 out of 370 PABPs (using Entrez Gene identifiers) that was not found in the compiled network. Because the available yeast interactome is much more complete, we have additionally mapped the human PABPs to yeast and performed the same analyses for the yeast homologs of the human PABPs. Here the yeast homologs for PABP data sets also participate in a larger number of physical interactions in the yeast interactome (supplemental Table 3 and supplemental Fig. 4B). In particular, they also show a clear tendency to form strongly connected sub-networks (supplemental Table 3 and Fig. 3B). This is in good agreement with our previous study of yeast PABPs (24).

Comparing Human and Yeast PABPs—A particularly important question in the analysis of PABPs is whether their functions are conserved between yeast and human. We have conducted similar protein array experiments for yeast and human and identified 529 and 397 PABPs in each organism, respectively. We subsequently asked if PABPs shared similar biological functions in these two quite different organisms.

Identifying Orthologs—Genes that have directly evolved from a single gene in a common ancestor and perform similar biological functions in two different species are known as orthologs (39). To identify orthologs of PABPs in yeast and human, we utilized the Inparanoid data base. The result, which can be found in supplemental Table 4, is the presence of 37 sets of orthologs in different functional categories, including but not limited to kinases, DNA- and RNA-binding proteins, metabolic enzymes, and ribosomal proteins. These results are also consistent with the domain search in yeast and human proteins.

We have identified four PABPs in yeast with unknown function, namely YPL150W, YKR030W, YNL175C, and YCR016W, and we tried to correlate these with their possible human orthologs. The human orthologs of these proteins are MARKL1, UNCL, a member of the RBM34 family, and finally a hypothetical protein (MGC11257). MARKL1 is a mitogen-associated protein/microtubule affinity-regulating kinase-like 1, involved in epithelial cell polarization processes and the control of cell division. The yeast ortholog, YPL150W, has kinase activity of unknown function. It is possible that YPL150W has a similar function to that of MARKL1 in human. UNCL is a protein with homology to unc-50 in *Caenorhabditis elegans* and is suggested to be an inner nuclear membrane protein that is associated with RNA and possibly plays a role in neural nicotinic receptor expression (50). The yeast ortholog of this protein is YKR030W, a Golgi membrane protein with unknown function that is expressed in yeast strains lacking nicotinic receptors. This protein, however, is suggested to be involved in different functions involving either cell wall synthesis or protein-vacuolar targeting (51, 52). An RNA-binding protein that belongs to the RBM34 family was identified to be an ortholog of YNL175C, a yeast PABP. The latter also contains an RNA-binding motif. They are both nuclear proteins and may play a role in the processing of rRNAs. Because they are both PABPs, such associa-

TABLE 4

The number of PABPs and domains present in both yeast and human

Pfam_id	Domain name	Function group	Proteins (human)	Domains (human)	Proteins (yeast)	Domains (yeast)
Pkinase	Protein kinase domain	Kinase	48	50	7	7
Pkinase_Tyr	Protein tyrosine kinase	Protein-tyrosine kinase activity	48	50		
RRM_1	RNA recognition motif	RNA binding	15	23	6	10
SH3_1	Src homology 3 domain	Signal transduction	9	9	1	3
Helicase_C	Helicase conserved C-terminal domain	Couples an ATPase activity to RNA binding and unwinding	8	8	9	9
DEAD	DEAD/DEAH box helicase	RNA metabolism	8	8	8	8
SH3_2	Variant SH3 domain	Signal transduction	8	8		
WD40	WD domain, G- β repeat	Protein interactions	7	22	8	39
ResIII	Type III restriction enzyme, res subunit	ATP binding	6	6		
zf-C2H2	Zinc finger, C2H2 type	Nucleic acid binding	6	39	3	7
PH	Pleckstrin homology domain	Intracellular signaling	6	7	3	3
MMR_HSR1	GTPase of unknown function	GTP binding	6	6	1	1
C1_3	C1-like domain	Protein kinase C	5	6		
APH	Phosphotransferase enzyme family	Protein kinase superfamily	5	5		
KH_1	KH domain	Nucleic acid binding	5	7	2	10
DnaJ	DnaJ domain	Chaperone (protein folding) system	2	2	5	5
AAA	ATPase family associated with various cellular activities	Chaperone-like functions that assist in the assembly, operation, or disassembly of protein complexes	1	1	6	6
TYA	TYA transposon protein	Transposition			8	8
UCH	Ubiquitin C-terminal hydrolase	Peptide bond hydrolysis			6	6
Mito_carr	Mitochondrial carrier protein	Energy transfer			5	15

tions with RNA are reasonably expected. The human hypothetical protein (MGC11257) is a homolog of human CYP2W1 (cytochrome P450, family 2, subfamily W, and polypeptide 1), which is involved in many metabolic reactions. We identified a yeast ortholog of this protein that is localized to the nucleolus and nucleus. Considering that most of the human cytochrome P450 family is localized to liver, drawing any conclusions concerning a functional orthology of these proteins is difficult at present.

Homolog Search—The result of a homolog search found that 194 human PABPs have 148 homologs of yeast PABPs. This number of homologs was statistically larger than the number of homologs obtained from a random bootstrap analysis for both human and yeast, which were 125 ± 10 and 115 ± 10 , respectively. This suggests that PABPs may in many cases be functionally similar in human and yeast. This homolog analysis demonstrates that PABPs from human and yeast are present in greater numbers than in the randomly selected sets. The abundance of such homologs suggests an enrichment in certain required core functions in both species.

Domain Search—To investigate whether a well studied common theme or a new unidentified protein family was present in the PABPs in yeast and humans, a domain search was conducted. It has been observed that recurring domains or conserved common domains are often observed in different species that take on similar cellular functions. The summary result of a domain search in both species is presented in Table 4. Overall, there are 75 common domains between yeast and human. The complete list of these domains can be found in supplemental Table 5. There are 402 and 294 distinct Pfam domains in yeast and human PABPs, respectively. Some of these domains are found multiple times with the overall number of domain occurrences of 655 and 658 for yeast and human, respectively. This suggests that PABPs are potentially able to interact with several cellular polyanions simultaneously. Not unexpectedly, the major domains present are involved in binding polyanions,

such as RNA or GTP. Interestingly, however, there are more human proteins containing signaling and cellular communication domains than in yeast. This could indicate more complex signal processing and regulatory controls, which are increasingly more specialized in a multicellular organism compared with the unicellular yeast. It can be seen that many protein domains identified by this study are consistent with the above mentioned orthologs study. It should also be emphasized that these analyses are limited to the domains that are represented in the population of human proteins on the array. Furthermore, many protein domains identified by this analysis are consistent with the previous orthological investigation.

Conclusions—Our approach toward identifying whether PABPs form some sort of functional network within cells was to initially identify these proteins by utilizing protein arrays probed with five typical cellular polyanions and subsequently building an interaction network among the identified proteins based on known cellular pathways. Rual *et al.* (53) and Stelzl *et al.* (54) have recently generated a draft of the human protein-protein interaction network with some overlap between the two data sets. Our data, however, focus on PABPs in humans in the hope of elucidating a new perspective concerning intracellular organization and the effect of such architecture on protein-protein interactions. One major drawback in current proteomic studies is the lack of a complete data base and that is seen here as well. For example, we were able to find only $\frac{2}{3}$ of our human PABPs among proteins currently recognized to be involved in protein-protein interactions. Considering the complicated nature of such interactions in human cells, this is not surprising. It should also be emphasized that the protein arrays we have employed include only a small fraction of the human proteome. Furthermore, the highly concentrated nature of the cellular interior would be expected to result in many more protein-polyanion interactions than those seen in these simple dilute experiments.

Polyanion-binding Protein Networks in the Human Proteome

An attractive hypothesis is that the polyanionic surfaces serve as an organizing matrix where protein-protein interactions occur. As an example, it was recently demonstrated that heparin and HS in the extracellular matrix alter the conformation of fibronectin causing it to interact with vascular endothelial growth factor-binding sites (55). This may represent polyanions serving as a type of interaction surface, but it will be necessary to search further for such examples if a more general phenomenon is involved. Similarly, the apparently enhanced interaction between polyanions and less structured proteins is consistent with a role for polyanionic surfaces in providing a stabilizing, chaperone-like environment.

A recent paper (56) has described the dynamics of protein assembly for the N-terminal domain of enzyme I and the phosphocarrier protein HPr of a bacterial phosphotransferase system as well as several other protein complexes. It demonstrated that not only nonspecific, transient collisions between the proteins play an important role in complex formation, but the charge distribution outside of the direct interaction surfaces may also modulate the rate of such associations to form transient encounter complexes of functional significance. From the point of view of the polyanion hypothesis, such weak complexes may often be of the polyanion/polycation type providing another important role for such interactions.

In this study, we have also demonstrated that many positively charged regions in PABPs are found in kinases and serve as kinase substrates. The presence of many other conserved domains among PABPs is also consistent with some type of critical function. Furthermore, it appears that PABPs seem to form some type of network with other human proteins. This is based on the observation that a large number of protein-protein interactions involve PABPs. A direct experimental demonstration of such a network, however, requires further validation.

We have previously proposed the crude analogy that if we consider the cell a city, then polyanions may behave like the roads, buildings, elevators, and other routes of transportation and sites of habitation, whereas the PABPs correspond to the individuals who interact and perform the various activities of the city (24). In the light of these new studies, however, a better analogy might be that a cell may manifest somewhat less organized behavior but rather be better compared with a tropical forest of less specific interactions between PABPs and polyanionic surfaces in which a dynamic set of interactions (which alters according to the requirements of a cell) govern the cellular behavior. All of the result files are available from the author.

Acknowledgments—We thank Richard Himes and Fernando Ausar. We also thank Aaron Smalter for his critical role in programming activities. We acknowledge Kim Bure, Clea D'sa, and Kevin Clancy at Invitrogen for their helpful discussions. Maria Miteva is acknowledged for help with PCE calculations.

REFERENCES

- Medalia, O., Weber, I., Frangakis, A. S., Nicastro, D., Gerisch, G., and Baumeister, W. (2002) *Science* **298**, 1209–1213
- Jones, L. S., Yazzie, B., and Middaugh, C. R. (2004) *Mol. Cell. Proteomics* **3**, 746–769
- Spitzer, J. J., and Poolman, B. (2005) *Trends Biochem. Sci.* **30**, 536–541
- Kurakin, A. (2005) *Dev. Genes Evol.* **215**, 46–52
- Ellis, R. J. (2001) *Trends Biochem. Sci.* **26**, 597–604
- Minton, A. P. (2001) *J. Biol. Chem.* **276**, 10577–10580
- Verkman, A. S. (2002) *Trends Biochem. Sci.* **27**, 27–33
- Conrad, H. E. (1998) *Heparin-binding Proteins*, Academic Press, San Diego
- Dabora, J. M., Sanyal, G., and Middaugh, C. R. (1991) *J. Biol. Chem.* **266**, 23637–23640
- Volkin, D. B., Tsai, P. K., Dabora, J. M., Gress, J. O., Burke, C. J., Linhardt, R. J., and Middaugh, C. R. (1993) *Arch. Biochem. Biophys.* **300**, 30–41
- Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B., and Blundell, T. L. (2000) *Nature* **407**, 1029–1034
- Tumova, S., Woods, A., and Couchman, J. R. (2000) *Int. J. Biochem. Cell Biol.* **32**, 269–288
- Manna, T., Sarkar, T., Poddar, A., Roychowdhury, M., Das, K. P., and Bhattacharyya, B. (2001) *J. Biol. Chem.* **276**, 39742–39747
- Rentzperis, D., Jonsson, T., and Sauer, R. T. (1999) *Nat. Struct. Biol.* **6**, 569–573
- Mach, H., Ryan, J. A., Burke, C. J., Volkin, D. B., and Middaugh, C. R. (1993) *Biochemistry* **32**, 7703–7711
- Tsai, P. K., Volkin, D. B., Dabora, J. M., Thompson, K. C., Bruner, M. W., Gress, J. O., Matuszewska, B., Keogan, M., Bondi, J. V., and Middaugh, C. R. (1993) *Pharm. Res. (N. Y.)* **10**, 649–659
- Kueltzo, L. A., and Middaugh, C. R. (2000) *Exp. Opin. Investig. Drugs* **9**, 2039–2050
- Kueltzo, L. A., Normand, N., O'Hare, P., and Middaugh, C. R. (2000) *J. Biol. Chem.* **275**, 33213–33221
- Wiedlocha, A., Madhus, I. H., Mach, H., Middaugh, C. R., and Olsnes, S. (1992) *EMBO J.* **11**, 4835–4842
- Brimacombe, D. B., Bennett, A. D., Wusteman, F. S., Gill, A. C., Dann, J. C., and Bostock, C. J. (1999) *Biochem. J.* **342**, 605–613
- Deleault, N. R., Geoghegan, J. C., Nishina, K., Kascak, R., Williamson, R. A., and Supattapone, S. (2005) *J. Biol. Chem.* **280**, 26873–26879
- Hasegawa, M., Crowther, R. A., Jakes, R., and Goedert, M. (1997) *J. Biol. Chem.* **272**, 33118–33124
- Li, J. P., Galvis, M. L., Gong, F., Zhang, X., Zcharia, E., Metzger, S., Vlodavsky, I., Kisilevsky, R., and Lindahl, U. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6473–6477
- Salamat-Miller, N., Fang, J., Seidel, C. W., Smalter, A. M., Assenov, Y., Albrecht, M., and Middaugh, C. R. (2006) *Mol. Cell. Proteomics* **5**, 2263–2278
- Yu, Q., and Toole, B. P. (1995) *BioTechniques* **19**, 122–129
- Apgar, J., Tseng, Y., Fedorov, E., Herwig, M. B., Almo, S. C., and Wirtz, D. (2000) *Biophys. J.* **79**, 1095–1106
- Templeton, D. M. (1988) *Int. J. Biol. Macromol.* **10**, 131–136
- Wollin, A., and Jaques, L. B. (1973) *Thromb. Res.* **2**, 377–382
- Schweitzer, B., Predki, P., and Snyder, M. (2003) *Proteomics* **3**, 2190–2199
- Miteva, M. A., Tuffery, P., and Villoutreix, B. O. (2005) *Nucleic Acids Res.* **33**, W372–W375
- Fang, J., Haas, R. J., Dong, Y., and Lushington, G. H. (2005) *BMC Bioinformatics* **6**, 277–284
- Bailey, T. L., and Elkan, C. (1994) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36
- Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Pagni, M., Ponting, C. P., Quevillon, E., Selengut, J., Sigrist, C. J., Silventoinen, V., Studholme, D. J., Vaughan, R., and Wu, C. H. (2005) *Nucleic Acids Res.* **33**, D201–D205
- Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) *Science* **298**, 1912–1934
- Diella, F., Cameron, S., Gemund, C., Linding, R., Via, A., Kuster, B., Sicheritz-Ponten, T., Blom, N., and Gibson, T. J. (2004) *BMC Bioinformatics* **5**, 79
- Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003) *Structure* **11**, 1453–1459

37. Albert, R. (2005) *J. Cell Sci.* **118**, 4947–4957
38. O'Brien, K. P., Remm, M., and Sonnhhammer, E. L. (2005) *Nucleic Acids Res.* **33**, D476–D480
39. Remm, M., Storm, C. E., and Sonnhhammer, E. L. (2001) *J. Mol. Biol.* **314**, 1041–1052
40. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhhammer, E. L. (2000) *Nucleic Acids Res.* **28**, 263–266
41. Hileman, R. E., Fromm, J. R., Weiler, J. M., and Linhardt, R. J. (1998) *BioEssays* **20**, 156–167
42. Sedlak, E., and Antalik, M. (1998) *Biopolymers* **46**, 145–154
43. Yeung, T., Terebiznik, M., Yu, L., Silviu, J., Abidi, W. M., Philips, M., Levine, T., Kapus, A., and Grinstein, S. (2006) *Science* **313**, 347–351
44. Pokholok, D. K., Zeitlinger, J., Hannett, N. M., Reynolds, D. B., and Young, R. A. (2006) *Science* **313**, 533–536
45. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) *Biochemistry* **41**, 6573–6582
46. Wright, P. E., and Dyson, H. J. (1999) *J. Mol. Biol.* **293**, 321–331
47. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002) *J. Mol. Biol.* **323**, 573–584
48. Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) *Biochemistry* **45**, 6873–6888
49. Dedmon, M. M., Patel, C. N., Young, G. B., and Pielak, G. J. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12681–12684
50. Fitzgerald, J., Kennedy, D., Viseshakul, N., Cohen, B. N., Mattick, J., Bateman, J. F., and Forsayeth, J. R. (2000) *Brain Res.* **877**, 110–123
51. Chantalat, S., Courbeyrette, R., Senic-Matuglia, F., Jackson, C. L., Goud, B., and Peyroche, A. (2003) *Mol. Biol. Cell* **14**, 2357–2371
52. Samanta, M. P., and Liang, S. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12579–12583
53. Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P., and Vidal, M. (2005) *Nature* **437**, 1173–1178
54. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlauff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. (2005) *Cell* **122**, 957–968
55. Mitsi, M., Hong, Z., Costello, C. E., and Nugent, M. A. (2006) *Biochemistry* **45**, 10319–10328
56. Tang, C., Iwahara, J., and Clore, M. (2006) *Nature* **444**, 383–386

Supplementary Material

A Network-Based Analysis of Polyanion Binding Proteins Utilizing Human Protein Arrays

Nazila Salamat-Miller¹, Jianwen Fang², Christopher Seidel³, Yassen Assenov⁴, Mario Albrecht⁴, and C. Russell Middaugh^{1,*}

¹Department of Pharmaceutical Chemistry, University of Kansas, 2030 Becker Dr., Lawrence, KS 66047.

² Bioinformatics Core Facility, University of Kansas, Lawrence, KS 66047.

³ Stowers Institute for Medical Research, 1000 E. 50th St., Kansas City, MO 64110.

⁴ Max Planck Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany

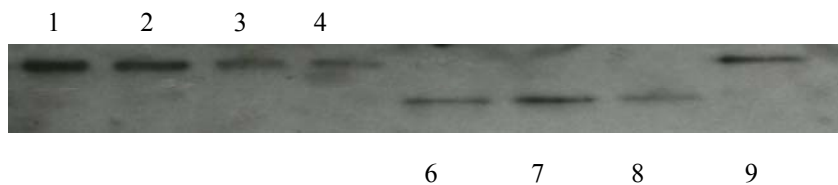
*Corresponding author: C. Russell Middaugh, Phone: 875-864-5813, Fax: 875-864-5814, e-mail: middaugh@ku.edu

Supplementary Figures

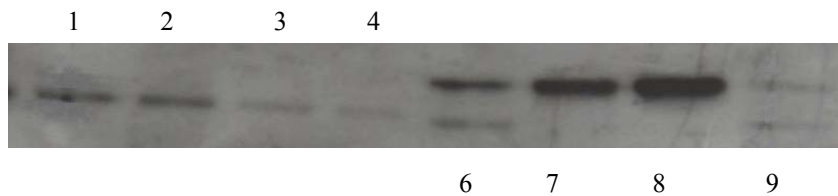
SF 1

Western blots of biotinylated A) actin and B) tubulin as well as dot-blots of C) heparin and HS, and D) DNA. Lines 1 to 4 in panels A and B represent biotinylated standards (200-12.5 fmole biotin) and lines 6-8 represent biotinylated proteins with a 3/1, 9/1, and 27/1 ratio of biotin to protein. The last line (9) represents the biotinylated BSA standard at a ratio of 9/1. This line in panel B is not as clearly visible; we suspect a low quality of the provided biotinylated BSA for this particular batch.

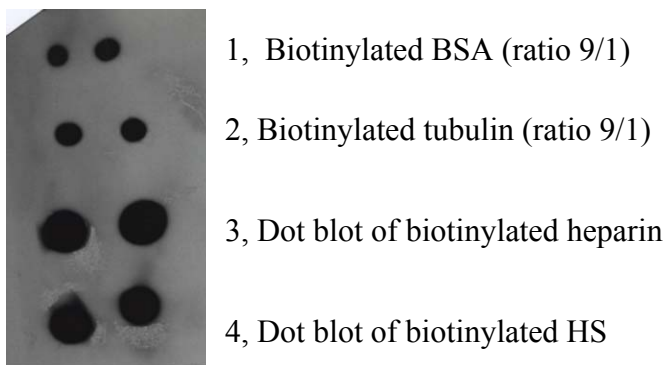
A



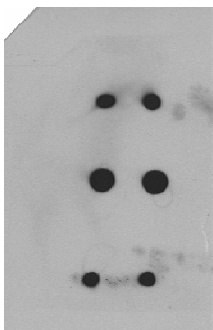
B



C



D



1, Biotinylated BSA (ratio 9/1)

2, Dot blot of biotinylated DNA

3, Biotinylated tubulin (ratio 9/1)

SF 2

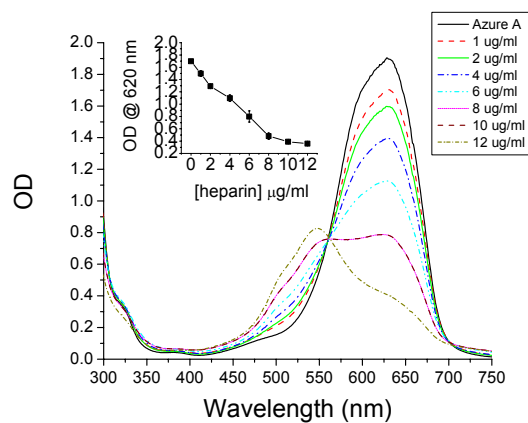
Similar results of this procedure have been reported previously (Salamat-Miller et al., In Press). Azure A, in the absence of heparin or HS, exhibited an absorption peak maximum at approximately 630 nm. Addition of heparin to the dye solution decreased the intensity of this peak and produced a new peak at a lower wavelength around 530 nm (characteristic of substances which produce a complex with this dye and demonstration of metachromacy, SF 2A and B) (Jaques, 1979). This assay was linear only at low concentrations of heparin since excess heparin interfered with color formation of the complex (Jaques, 1947). Although solutions of HS and Azure A did not produce a new peak at 530 nm, the intensity of the peak maximum of the dye was reduced following the addition of HS to the dye solution in a concentration dependent manner (SF 2C and 2D).

Azure A, similar to other thiazine dyes, is believed to form dimers through stacking interactions. It has been shown that certain structural arrangements of polysaccharides produce charge densities and geometries that permit the dimer to form a complex with the polysaccharide. This is detectable at lower wavelengths when compared to the wavelength of the original dye (Wollin, 1973). This band was observed for heparin-dye interactions at 530 nm and is shown in the supplementary Figures (SF) 2A and 2B. Azure A in the presence of HS did not manifest this new, single band but demonstrated a decrease in its peak maximum intensity in a HS concentration dependent fashion (SF 2C and 2D). This fact and the change in the color of the dye solution upon addition of HS were both indicative of complex formation. Sample heterogeneity has been shown to suppress the formation of a new band when compared with more homogenous solutions (Templeton, 1988). Thus, we speculate that the HS solutions

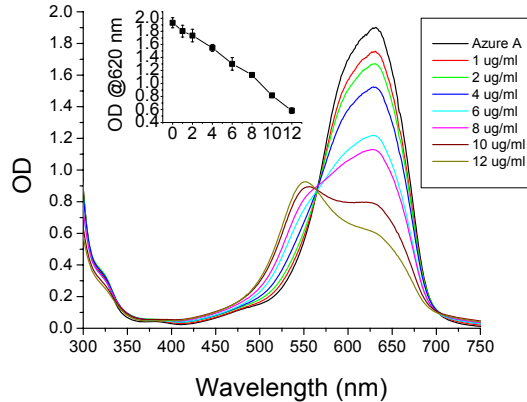
used here may be more heterogeneous than heparin. Additionally, the lower charge density of HS may play a role in this phenomenon (Wollin, 1973).

The ultraviolet absorption spectra are shown for A) heparin, B) biotinylated heparin, C) HS, and D) biotinylated HS in presence of azure A. The experiments were conducted in triplicate.

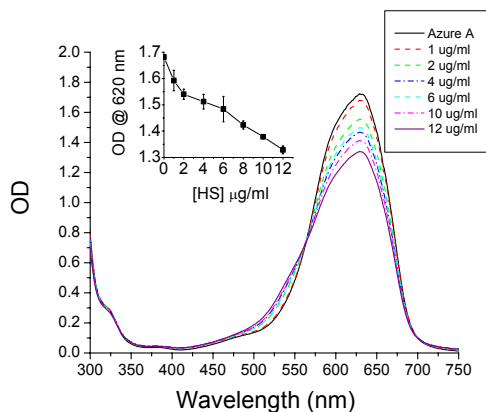
A)



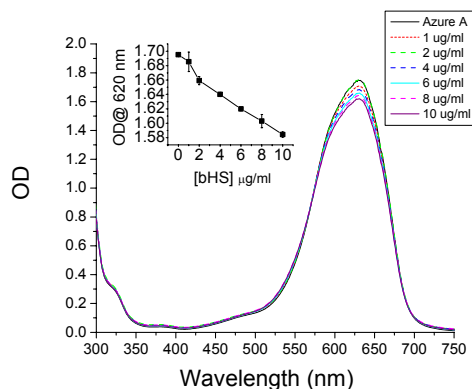
B)



C)



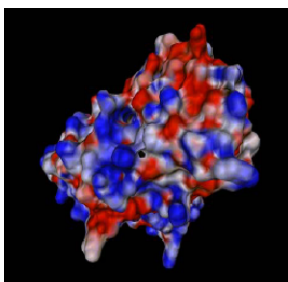
D)



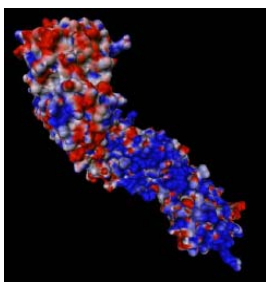
SF 3

Calculations of surface electrochemical potential were conducted using the protein continuum electrostatics (PCE) web tool <http://bioserv.rpbs.jussieu.fr/PCE> with the default values of 4.0, 80.0, and 0.1 for protein internal and solvent dielectric constants, and ionic strength, respectively. Surface potential calculations were not possible for E3 ubiquitin-proteins lieges CBL-B (gene symbol: CBLB, PDB code: 2BZ8) due to the missing backbone atom for residue Lys58A: “CA”. The same was true for IIMO (gene symbol: LIG3) due to the quality of the X-ray structure which caused atoms to collide and, therefore, prevent the surface potential computation from reaching completion (personal communication with the program author). The following figures are named with the gene symbol of the proteins following by their PDB code. Each figure is rotated and the best representation of the presumptive polyanion binding site(s) is presented. In some structures, the proposed polyanion binding sites are more pronounced than others (for example in 1MJ D and 1I0Z compared to 1T0P).

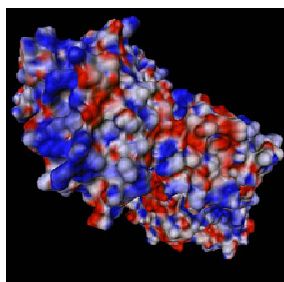
ANXA2, 1 W7B



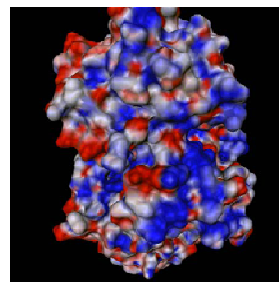
APC, 1JPP



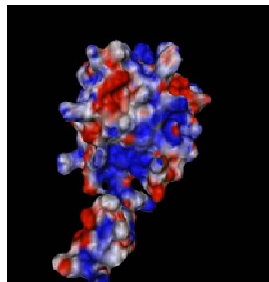
BRAF, 1UWH



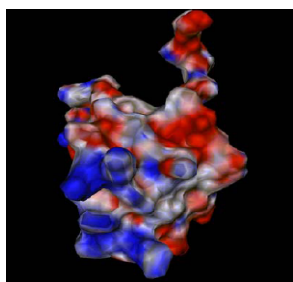
BTK, 1K2P



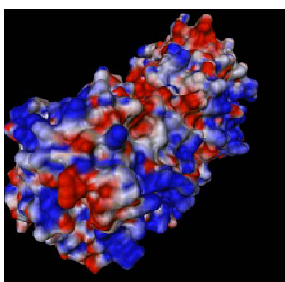
DCX, 1MJD



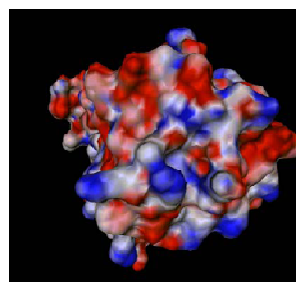
EPS8L3, 1WXT



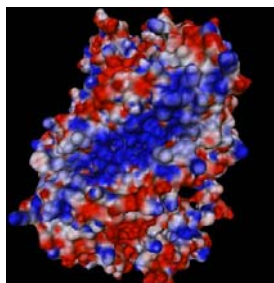
F2, 2AFQ



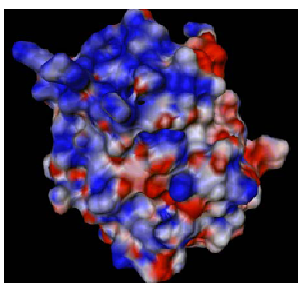
ICAM 3, 1T0P



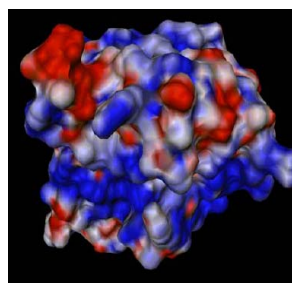
LDHB, 1I0Z



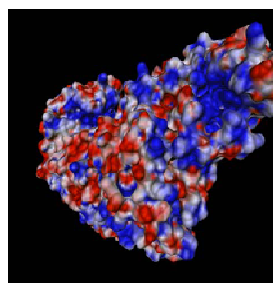
MGMT, 1EH7



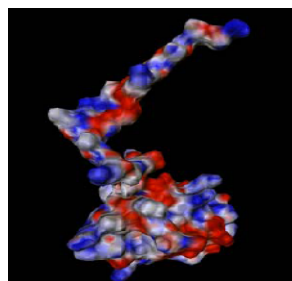
MPG, 1BNK



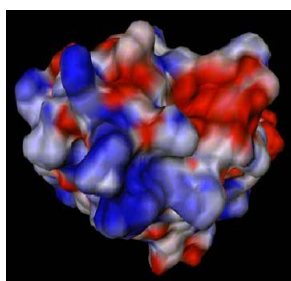
PAPSS2, 2AX4



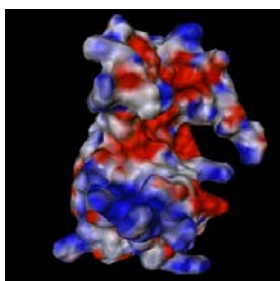
PCCA, 2CQY



PNKP, 2BRF



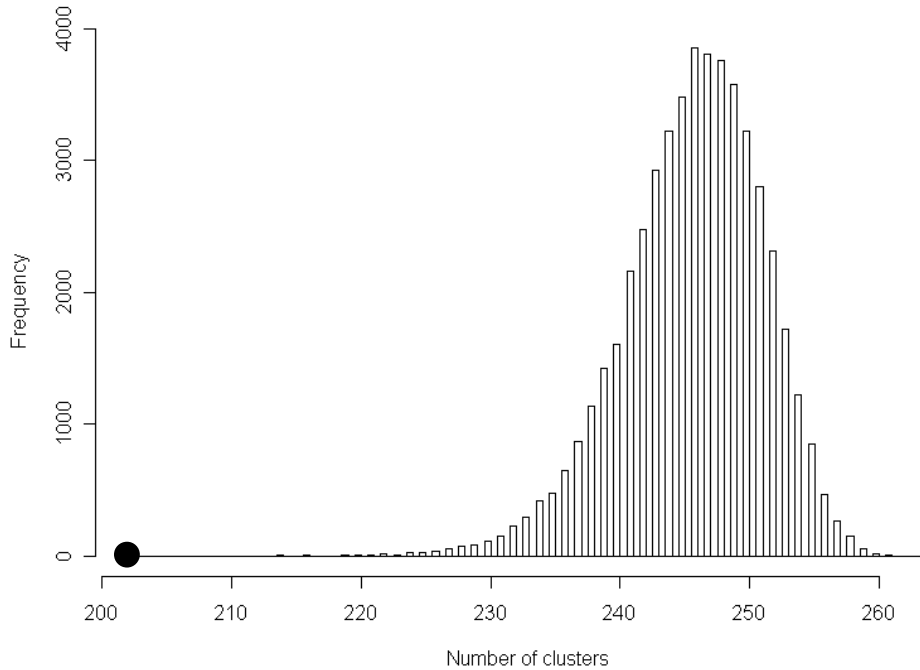
PBM23, 2CQ4



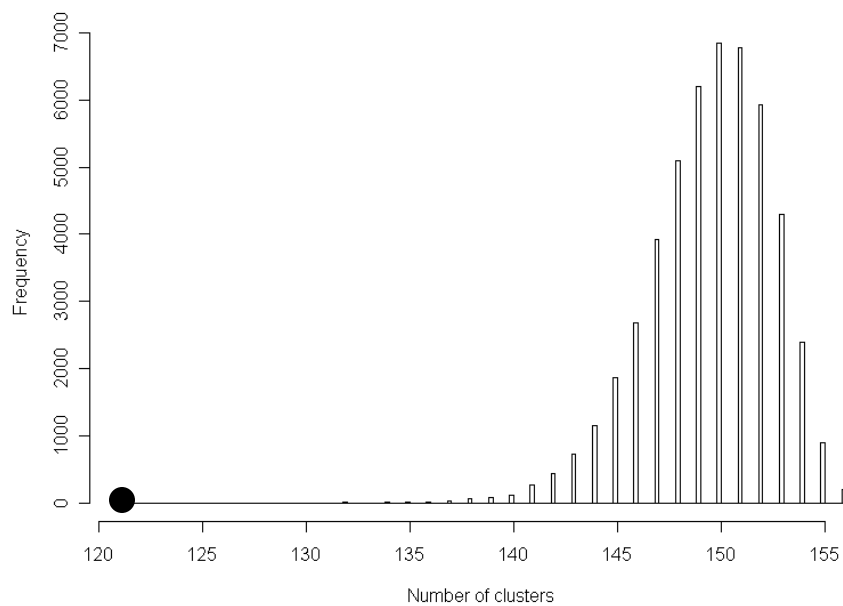
SF 4

A) Histograms for the number of clusters with interacting proteins in random human sub-networks. The horizontal axis of the histogram represents the different values obtained for the number of clusters in the generated sub-networks. The vertical axis shows how many random sub-networks contained the given number of clusters. These histograms are produced for three different statistical levels (+1SD, +2SD and +3SD) of PABP detection to estimate the p -values shown in Table 3 in the text. The black circle represents the number of clusters in the sub-network formed by PABPs only. It can be seen that the number of clusters for PABPs are fewer than the generated respective random sets. This indicates a higher connectivity for the PABPs compared to more random clusters which manifest lack of connectivity between the proteins in each cluster.

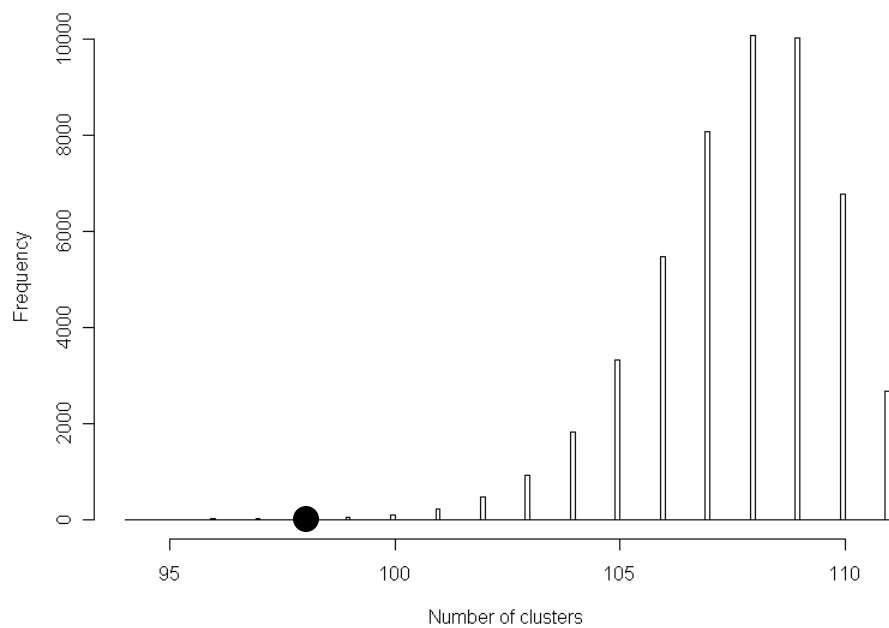
a) 1SD



b) 2SD

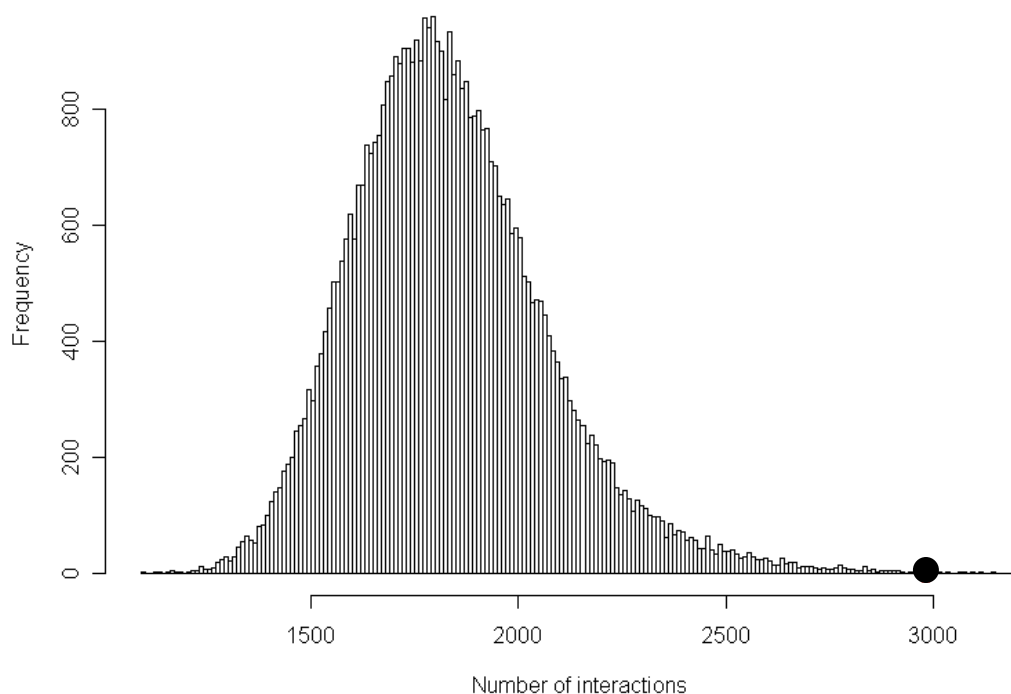


c) 3SD

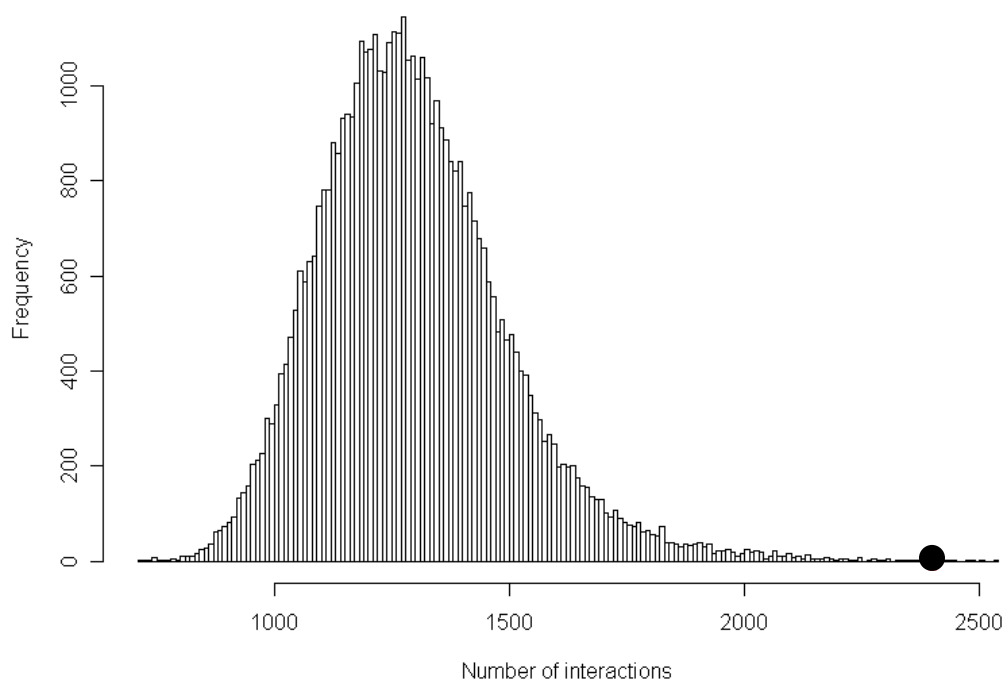


B) Histograms for the number of protein-protein interactions involving yeast proteins that are homologous to human PABPs in random yeast sub-networks. The histograms are produced for three different statistical levels +1SD, +2SD and +3SD of PABP detection to estimate the *p*-values shown in ST 3. The black circle represents the number of interactions in the sub-network formed by PABP homologues.

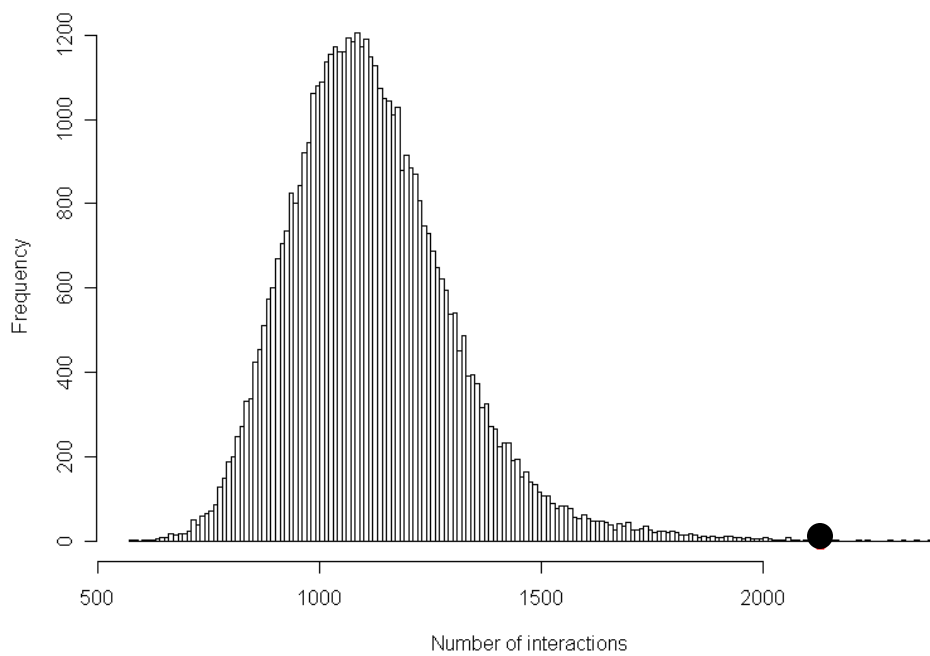
a) 1SD



b) 2SD



c) 3SD



Supplementary Tables

ST 1

The standard ORF list of PABPs generated using human protein arrays. The reported data contain PABPs with low, medium, and high interaction confidence levels.

Actin	Tubulin	Heparin	HS	DNA
BC000381	BC000381	BC000043	BC000381	BC000381
BC000442	BC000442	BC000807	BC000953	BC000953
BC000914	BC000897	BC001280	BC001088	BC001088
BC000953	BC001088	BC001709	BC002559	BC002559
BC002559	BC001662	BC002559	BC002733	BC002733
BC002582	BC005004	BC002606	BC005004	BC005004
BC005004	BC006376	BC002880	BC006376	BC006376
BC007347	BC007888	BC003132	BC007888	BC007888
BC007888	BC008623	BC003168	BC008623	BC008623
BC007957	BC009348	BC004292	BC009348	BC009348
BC008077	BC009762	BC004514	BC009623	BC009623
BC008623	BC009820	BC005004	BC009762	BC009762
BC008825	BC010697	BC005955	BC010467	BC010467
BC009046	BC010947	BC007228	BC010501	BC010501
BC009348	BC012493	BC007888	BC010697	BC010697
BC009650	BC013966	BC008058	BC011842	BC011842
BC009762	BC014949	BC008077	BC012576	BC012576
BC010357	BC019598	BC008623	BC013051	BC013051
BC010451	BC020981	BC009348	BC013966	BC013966
BC010467	BC028040	BC009564	BC014298	BC014298
BC010501	BC028218	BC009623	BC016778	BC016778
BC010697	BC028301	BC009762	BC016967	BC016967
BC010947	BC029378	BC010357	BC017212	BC017212
BC011668	BC029611	BC010467	BC018823	BC018823
BC011842	BC031682	BC011668	BC020647	BC020647
BC011885	BC032485	BC011842	BC020981	BC020981
BC012576	BC032852	BC012472	BC024187	BC024187
BC013051	BC033758	BC012576	BC025345	BC025345
BC013073	BC033856	BC012865	BC027607	BC027607
BC013966	BC038976	BC013073	BC028040	BC028040
BC014298	BC045535	BC014298	BC028301	BC028301
BC014949	BC051885	BC014949	BC028711	BC028711
BC014959	BC053557	BC014959	BC029378	BC029378
BC016276	BC053595	BC014991	BC029611	BC029611
BC016967	BC056415	BC015109	BC031228	BC031228
BC019598	BC063275	BC015505	BC031682	BC031682
BC020647	BC067735	BC015749	BC032347	BC032347

BC020981	BC067755	BC016609	BC032654	BC032654
BC022344	NM_000282	BC016842	BC032852	BC032852
BC022361	NM_000989	BC017212	BC033621	BC033621
BC022524	NM_000993	BC018823	BC033758	BC033758
BC025266	NM_001018	BC019598	BC033856	BC033856
BC025404	NM_001022	BC020647	BC035314	BC035314
BC026317	NM_001023	BC020981	BC038808	BC038808
BC027607	NM_001697	BC026317	BC038976	BC038976
BC028040	NM_003160	BC026961	BC043619	BC043619
BC028151	NM_003582	BC028040	BC045535	BC045535
BC028218	NM_003600	BC028244	BC051885	BC051885
BC029378	NM_003621	BC028301	BC053557	BC053557
BC029611	NM_003897	BC028954	BC063500	BC063500
BC029760	NM_004073	BC029378	BC066938	BC066938
BC031228	NM_004114	BC029611	BC067735	BC067735
BC031682	NM_004333	BC030020	BC067755	BC067755
BC032347	NM_004384	BC030219	NM_000723	NM_00072
BC032852	NM_004765	BC030590	NM_000970	NM_00097
BC033529	NM_005517	BC031010	NM_000984	NM_00098
BC033621	NM_006205	BC031549	NM_000989	NM_00098
BC033758	NM_006298	BC031682	NM_000993	NM_00099
BC033856	NM_012279	BC032347	NM_001002913	NM_00100
BC034954	NM_013375	BC032485	NM_001013	NM_00101
BC035314	NM_013401	BC032851	NM_001014	NM_00101
BC038105	NM_013975	BC032852	NM_001018	NM_00101
BC038808	NM_014062	BC033621	NM_001022	NM_00102
BC038976	NM_014720	BC033758	NM_001023	NM_00102
BC039711	NM_014868	BC033790	NM_001029	NM_00102
BC041769	NM_015014	BC033856	NM_001280	NM_00128
BC043619	NM_015584	BC034954	NM_001325	NM_00132
BC045535	NM_015640	BC035143	NM_001697	NM_00169
BC051885	NM_015939	BC035601	NM_002677	NM_00267
BC053557	NM_016304	BC038838	NM_002749	NM_00274
BC060773	NM_016355	BC040183	NM_002802	NM_00280
BC063500	NM_016483	BC043619	NM_002922	NM_00292
BC065525	NM_017588	BC044584	NM_003582	NM_003582
BC067735	NM_017838	BC045535	NM_003583	NM_003583
BC067755	NM_018956	BC047536	NM_003621	NM_003621
BC067773	NM_020166	BC049215	NM_003897	NM_003897
BC070073	NM_020444	BC052303	NM_004114	NM_004114
CTL2109	NM_021104	BC053365	NM_004214	NM_004214
NM_000061	NM_031412	BC054892	NM_004310	NM_004310
NM_000282	NM_031417	BC056415	NM_004966	NM_004966
NM_000989	NM_031465	BC063500	NM_005011	NM_005011
NM_000993	NM_032345	BC064945	NM_005517	NM_005517
NM_001002913	NM_032350	BC065525	NM_005720	NM_005720
NM_001014	NM_033642	BC066938	NM_006205	NM_006205
NM_001018	NM_138551	BC067773	NM_006298	NM_006298

NM_001022	NM_144602	NM_000061	NM_006528	NM_006528
NM_001023	NM_152376	NM_000723	NM_006607	NM_006607
NM_001029	NM_153207	NM_000970	NM_006779	NM_006779
NM_001280	NM_175923	NM_000989	NM_007162	NM_007162
NM_001501	NM_178151	NM_001002913	NM_012279	NM_012279
NM_001697		NM_001013	NM_013342	NM_013342
NM_001896		NM_001018	NM_013375	NM_013375
NM_002031		NM_001022	NM_013975	NM_013975
NM_002082		NM_001023	NM_014046	NM_014046
NM_002300		NM_001203	NM_014047	NM_014047
NM_002498		NM_001325	NM_014062	NM_014062
NM_002539		NM_001348	NM_014868	NM_014868
NM_002749		NM_001449	NM_015014	NM_015014
NM_003146		NM_001790	NM_015640	NM_015640
NM_003160		NM_001952	NM_015933	NM_015933
NM_003390		NM_002082	NM_015939	NM_015939
NM_003582		NM_002688	NM_016271	NM_016271
NM_003621		NM_002738	NM_016304	NM_016304
NM_003897		NM_002749	NM_016483	NM_016483
NM_004073		NM_002802	NM_016505	NM_016505
NM_004114		NM_003621	NM_017503	NM_017503
NM_004148		NM_003897	NM_017588	NM_017588
NM_004304		NM_003952	NM_017838	NM_017838
NM_004384		NM_004055	NM_018047	NM_018047
NM_004645		NM_004384	NM_018454	NM_018454
NM_004765		NM_004699	NM_020239	NM_020239
NM_004846		NM_004760	NM_020444	NM_020444
NM_004966		NM_004765	NM_021104	NM_021104
NM_005011		NM_004809	NM_021627	NM_021627
NM_005044		NM_005435	NM_021639	NM_021639
NM_005517		NM_005441	NM_022839	NM_022839
NM_005592		NM_005517	NM_024041	NM_024041
NM_005720		NM_005558	NM_024901	NM_024901
NM_006205		NM_005610	NM_031412	NM_031412
NM_006259		NM_006205	NM_031465	NM_031465
NM_006293		NM_006251	NM_031910	NM_031910
NM_006298		NM_006259	NM_032338	NM_032338
NM_006388		NM_006433	NM_032345	NM_032345
NM_006433		NM_006607	NM_032350	NM_032350
NM_006607		NM_006775	NM_032848	NM_032848
NM_006779		NM_006779	NM_033256	NM_033256
NM_007162		NM_007054	NM_033642	NM_033642
NM_012279		NM_007162	NM_080659	NM_080659
NM_013342		NM_007285	NM_130398	NM_130398
NM_013375		NM_014046	NM_138551	NM_138551
NM_013975		NM_014062	NM_138775	NM_138775
NM_014044		NM_014815	NM_144602	NM_144602
NM_014047		NM_015138	NM_144971	NM_144971

NM_014062		NM_015640	NM_144982	NM_144982
NM_014720		NM_015871	NM_152736	NM_152736
NM_014868		NM_015939	NM_153207	NM_153207
NM_015138		NM_016271	NM_153332	NM_153332
NM_015640		NM_016281	NM_175923	NM_175923
NM_015939		NM_016304	NM_178151	NM_178151
NM_016093		NM_016355	NM_199334	NM_199334
NM_016271		NM_016520	NM_206834	NM_206834
NM_016281		NM_016606	PM_2108	PM_2108
NM_016304		NM_016653	PM_2120	PM_2120
NM_016355		NM_017588		
NM_016457		NM_018454		
NM_016505		NM_018650		
NM_016653		NM_020234		
NM_017503		NM_020236		
NM_017588		NM_020239		
NM_017838		NM_020444		
NM_018025		NM_020954		
NM_018047		NM_021104		
NM_018107		NM_021117		
NM_018454		NM_021627		
NM_018457		NM_021639		
NM_018956		NM_022100		
NM_020166		NM_024749		
NM_020239		NM_024901		
NM_020444		NM_025061		
NM_020666		NM_031412		
NM_021104		NM_031417		
NM_021627		NM_031465		
NM_021639		NM_031992		
NM_021822		NM_032338		
NM_022839		NM_032345		
NM_024309		NM_032506		
NM_025061		NM_032693		
NM_031412		NM_032848		
NM_031417		NM_032906		
NM_031465		NM_032929		
NM_031910		NM_052845		
NM_032338		NM_080659		
NM_032345		NM_130398		
NM_032350		NM_138775		
NM_033642		NM_139209		
NM_052877		NM_144971		
NM_130398		NM_145269		
NM_133480		NM_152474		
NM_138334		NM_152688		
NM_138551		NM_152736		
NM_138779		NM_152770		

NM_139209		NM_152789		
NM_144595		NM_153207		
NM_144602		NM_173618		
NM_144982		NM_178151		
NM_145055		NM_182692		
NM_145063		NM_194326		
NM_145702		NM_199334		
NM_152332		NM_203284		
NM_152373		NM_203305		
NM_152376		NM_203350		
NM_152736		NM_206834		
NM_152789		PM_2108		
NM_153207		PM_2120		
NM_153332				
NM_153339				
NM_175923				
NM_182692				
NM_203454				
NP_005219				
XM_027237				
XM_027237				

Gene Symbol	PDB	Protein name	Notes
ANXA2	1W7B	Annexin A2	
APC	1JPP	Anaphase-promoting complex subunit 1	The Structure of a beta-Catenin Binding Repeat from Adenomatous Polyposis Coli (APC) in Complex with beta-Catenin
BRAF	1UWH	B-Raf proto-oncogene serine/threonine-protein kinase	Liganded to C ₂₁ H ₁₆ N ₄ O ₃ Cl F ₃
BTK	1K2P	Tyrosine-protein kinase BTK	
CBLB	2BZ8	E3 ubiquitin-protein ligase CBL-B	contains sodium ion
DCX	1MJD	Neuronal migration protein doublecortin	
EPS8L3	1WXT	Epidermal growth factor receptor kinase substrate 8-like protein 3	
F2	2AFQ	coagulation factor II (thrombin)	
ICAM3	1T0P	Intercellular adhesion molecule 3 [Precursor]	C ₈ H ₁₅ N O ₆ is the ligand
LDHB	1I0Z	L-lactate dehydrogenase B chain	Complexed with NADH and oxamate
LIG3	1IMO	ligase III, DNA, ATP-dependent Bottom of Form	
MGMT	1EH7	Methylated-DNA-protein-cysteine methyltransferase	Contains chemical component C ₄ H ₉ N O ₂ S
MPG	1BNK	N-methylpurine-DNA glycosylase Bottom of Form	Human 3-methyladenine DNA glycosylase complexed to DNA
PAPSS2	2AX4	Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthetase 2	Contains ADP
PCCA	2CQY	Propionyl-CoA carboxylase alpha chain, mitochondrial [Precursor]	Beta domain
PNKP	2BRF	Bifunctional polynucleotide phosphatase/kinase	FHA domain
RBM23	2CQ4	Probable RNA-binding protein 23	Solution structure of RNA binding domain in RNA binding motif protein 23

ST 3

This table lists the p -values estimated for the number of protein-protein interactions involving yeast proteins homologous to human PABPs (SF 4B). This also demonstrates the number of clusters for these proteins (Figure 3B in the text). The p -values show the significance of the corresponding sub-network for each statistical PABP detection level (defined by the standard deviation).

PABP type	SD ¹	Number of PABPs in human	Number of PABPs mapped to homologues	Number of homologues in yeast	Homologues present in the yeast interactome	PPI involving PABPs (p -value)	95% CI ²		Clusters of PABPs (p -value)	95% CI	
Actin-binding	1	202	100	99	98	0.0001	±	0.0001	0.00002	±	0.00004
Actin-binding	2	100	47	66	65	0.004	±	0.0006	0.00004	±	0.00006
Actin-binding	3	59	29	47	46	0.007	±	0.0007	0.00020	±	0.00012
DNA-binding	1	170	73	87	85	0.0001	±	0.0001	0.00002	±	0.00004
DNA-binding	2	90	37	54	53	0.0005	±	0.0002	0.00010	±	0.00009
DNA-binding	3	56	27	46	45	0.001	±	0.0002	0.00012	±	0.00010
Heparin-binding	1	194	90	111	110	0.001	±	0.0003	0.00002	±	0.00004
Heparin-binding	2	99	43	73	73	0.002	±	0.0004	0.00002	±	0.00004
Heparin-binding	3	59	25	57	57	0.002	±	0.0004	0.00002	±	0.00004
HS-binding	1	143	58	83	83	0.001	±	0.0002	0.00004	±	0.00006
HS-binding	2	92	37	52	52	0.002	±	0.0004	0.00002	±	0.00004
HS-binding	3	60	24	35	35	0.005	±	0.0006	0.00002	±	0.00004
Tubulin-binding	1	90	46	63	62	0.013	±	0.0010	0.00154	±	0.00034
Tubulin-binding	2	58	31	50	49	0.012	±	0.0010	0.00050	±	0.00020
Tubulin-binding	3	37	19	43	42	0.010	±	0.0009	0.00034	±	0.00016
All	1	397	192	148	145	0.0003	±	0.0002	0.00002	±	0.00004
All	2	227	105	104	102	0.0002	±	0.0001	0.00002	±	0.00004
All	3	156	72	90	88	0.0003	±	0.0002	0.00004	±	0.00006

1) SD: Standard Deviation

2) CI: Confidence Interval

ST 4

The output of Inparanoid for identification of protein orthologs in yeast and human; the underlined functions have been added to the table manually.

```
#####
37 clusters of orthologs
99 in-paralogs from FASTAFILE1
37 in-paralogs from FASTAFILE2
Grey zone 0 bits
Score cutoff 50 bits
In-paralogs with confidence less than 0.05 not shown
Sequence overlap cutoff 0.5
Group merging cutoff 0.5
Scoring matrix BLOSUM62
#####
```

Group of orthologs #1. Best score 637 bits

Score difference with first non-orthologous sequence - FASTAFILE1:637
FASTAFILE2:463

gi 4506206 ref NM_002802.	100.00%	YDL007W	100.00%
gi 4506206 ref NM_002802.	100.00%		

26S proteasome

ATPase subunit of proteasome

Group of orthologs #2. Best score 451 bits

Score difference with first non-orthologous sequence - FASTAFILE1:398
FASTAFILE2:244

gi 41327774 ref NM_016355	100.00%	YHR065C	100.00%
gi 41327774 ref NM_016355	100.00%		
gi 41327774 ref NM_016355	100.00%		
gi 41327774 ref NM_016355	100.00%		

DEAD box protein family (RNA helicase)

Ribosomal RNA processing

Group of orthologs #3. Best score 429 bits

Score difference with first non-orthologous sequence - FASTAFILE1:346
FASTAFILE2:332

gi 38708325 ref NM_001896	100.00%	YIL035C	100.00%
---------------------------	---------	---------	---------

gi 38708325 ref NM_001896	100.00%		
gi 4503096 ref NM_001896.	100.00%		
gi 4503096 ref NM_001896.	100.00%		
Casein Kinase 2 (CSNK2A2)		Proteins kinase CK2 alpha subunit	
gi 15079695 gb BC011668.1	59.80%		
gi 15079695 gb BC011668.1	59.80%		
gi 15079695 gb BC011668.1	59.80%		

Group of orthologs #4. Best score 411 bits
Score difference with first non-orthologous sequence - FASTAFILE1:411
FASTAFILE2:411

gi 4557832 ref NM_000282.	100.00%	YGL062W	100.00%
gi 4557832 ref NM_000282.	100.00%		
gi 4557832 ref NM_000282.	100.00%		
Propionyl coenzyme A carboxylase		Pyruvate carboxylase	
gi 13518227 ref NM_020166	100.00%		
gi 13518227 ref NM_020166	100.00%		
gi 13518227 ref NM_020166	100.00%		
methyl crotonoyl coenzyme A carboxylase 1			

Group of orthologs #5. Best score 393 bits
Score difference with first non-orthologous sequence - FASTAFILE1:233
FASTAFILE2:393

gi 33188448 ref NM_182692	100.00%	YMR216C	100.00%
gi 33188448 ref NM_182692	100.00%		
SFRS proteins kinase 2 (SRPK2)		SR protein kinase (mRNA metabolism)	
gi 7657628 ref NM_014370.	35.12%		

Group of orthologs #6. Best score 377 bits
Score difference with first non-orthologous sequence - FASTAFILE1:326
FASTAFILE2:130

gi 45219879 gb BC066938.1	100.00%	YNL112W	100.00%
gi 45219879 gb BC066938.1	100.00%		
DEAD box polypeptide 43		ATP-dependent RNA helicase of the DEAD-box protein family	
gi 15930130 gb BC015505.1	100.00%		

gi|15930130|gb|BC015505.1 100.00%

RNA-helicase-related protein

Group of orthologs #7. Best score 287 bits

**Score difference with first non-orthologous sequence - FASTAFILE1:287
FASTAFILE2:110**

gi|34190861|gb|BC030020.2 100.00% YMR290C 100.00%

DEAD box polypeptide 55

Helicase associated with Set1

Group of orthologs #8. Best score 281 bits

**Score difference with first non-orthologous sequence - FASTAFILE1:31
FASTAFILE2:281**

gi|8922711|ref|NM_018243. 100.00% YLR314C 100.00%

Hypothetical protein FLJ10849

Septin (required for Cytokinesis)

gi|21945063|ref|NM_144710 63.79%

Septin 10 (SEPT 10)

Group of orthologs #9. Best score 280 bits

**Score difference with first non-orthologous sequence - FASTAFILE1:280
FASTAFILE2:280**

gi|23271507|gb|BC016967.1 100.00% YOL066C 100.00%

gi|23271507|gb|BC016967.1 100.00%

Cl8B11 homolog

DRAP deaminase

Group of orthologs #10. Best score 223 bits

**Score difference with first non-orthologous sequence - FASTAFILE1:223
FASTAFILE2:97**

gi|5453689|ref|NM_006145. 100.00% YNL007C 100.00%

DnaJ (Hsp 40)

Slt 4 suppressor (type II Hsp 40)

Group of orthologs #11. Best score 205 bits

Score difference with first non-orthologous sequence - FASTAFILE1:40

FASTAFILE2:205

gi 13899224 ref NM_031417	100.00%	YPL150W	100.00%
gi 13899224 ref NM_031417	100.00%		
gi 13899224 ref NM_031417	100.00%		

**MAP/microtubule affinity-regulating
kinase like 1 (MARKL1)**

**Unknown protein with kinase
activity**

gi|8923921|ref|NM_018650. 25.35%

MARKL1

Group of orthologs #12. Best score 169 bits

Score difference with first non-orthologous sequence - FASTAFILE1:35

FASTAFILE2:169

gi 21618348 ref NM_145109	100.00%	YPL140C	100.00%
---------------------------	---------	---------	---------

**Mitogen-activated protein kinase
Kinase 3 (MAP2K3)**

Mitogen-activated kinase kinase

Group of orthologs #13. Best score 167 bits

Score difference with first non-orthologous sequence - FASTAFILE1:5

FASTAFILE2:167

gi 4557510 ref NM_001348.	100.00%	YDL101C	100.00%
---------------------------	---------	---------	---------

**Death-associated protein kinase 3
(DAPK3)**

**Protein kinase
(required for DNA damage-
induced transcription)**

Group of orthologs #14. Best score 166 bits

Score difference with first non-orthologous sequence - FASTAFILE1:44

FASTAFILE2:166

gi 20986496 ref NM_002749	100.00%	YMR139W	100.00%
gi 20986496 ref NM_002749	100.00%		
gi 20986496 ref NM_002749	100.00%		

**Mitogen-activated protein kinase 7
(DAPK7)**

Regulator of IME2 (a kinase)

Group of orthologs #15. Best score 159 bits

Score difference with first non-orthologous sequence - FASTAFILE1:159

FASTAFILE2:159

gi 37059764 ref NM_014044	100.00%	YKR030W	100.00%
---------------------------	---------	---------	---------

Unc-50 homolog (C.elegans)

Unknown protein in Glogi membrane

Group of orthologs #16. Best score 153 bits

Score difference with first non-orthologous sequence - FASTAFILE1:153

FASTAFILE2:153

gi 20987238 gb BC029760.1	100.00%	YHL013C	100.00%
---------------------------	---------	---------	---------

CGI-77 protein

ovarian tumor like super family (OTU)

domain containing 6B

Uncharacterized member of OTU

Group of orthologs #17. Best score 147 bits

Score difference with first non-orthologous sequence - FASTAFILE1:147

FASTAFILE2:147

gi 16753226 ref NM_000970	100.00%	YML073C	100.00%
---------------------------	---------	---------	---------

gi 16753226 ref NM_000970	100.00%
---------------------------	---------

Ribosomal protein L6 (RPL6)

Large subunit of ribosomal protein
L6A (L17A) (rpl8) (YL16)

Group of orthologs #18. Best score 145 bits

Score difference with first non-orthologous sequence - FASTAFILE1:70

FASTAFILE2:53

gi 16554626 ref NM_017588	100.00%	YCR072C	100.00%
---------------------------	---------	---------	---------

gi 16554626 ref NM_017588	100.00%
---------------------------	---------

gi 16554626 ref NM_017588	100.00%
---------------------------	---------

gi 16554626 ref NM_017588	100.00%
---------------------------	---------

gi 16554626 ref NM_017588	100.00%
---------------------------	---------

WD-repeat domain 5 (WDR5)

Ribosome assembly WD-repeat protein
involved in ribosomal biogenesis

Group of orthologs #19. Best score 138 bits

Score difference with first non-orthologous sequence - FASTAFILE1:138

FASTAFILE2:138

gi 6857803 ref NM_007054.	100.00%	YKL079W	100.00%
Kinesin family member 3A (KIF3A)		Suppressor of MYO2-66 (N-terminal domain is related to the motor domain kinesins)	

Group of orthologs #20. Best score 134 bits

Score difference with first non-orthologous sequence - FASTAFILE1:134

FASTAFILE2:134

gi 14591915 ref NM_001023	100.00%	YHL015W	100.00%
gi 14591915 ref NM_001023	100.00%		
gi 14591915 ref NM_001023	100.00%		
gi 14591915 ref NM_001023	100.00%		
Ribosome protein 20 (RPS20)		Ribosome protein S20	

Group of orthologs #21. Best score 113 bits

Score difference with first non-orthologous sequence - FASTAFILE1:113

FASTAFILE2:113

gi 7661531 ref NM_014062.	100.00%	YOR056C	100.00%
gi 7661531 ref NM_014062.	100.00%		
gi 7661531 ref NM_014062.	100.00%		
gi 7661531 ref NM_014062.	100.00%		
gi 7661531 ref NM_014062.	100.00%		
Ortholog of mouse nin one binding protein (NOB1P)		Nin1 binding protein	

Group of orthologs #22. Best score 112 bits

Score difference with first non-orthologous sequence - FASTAFILE1:112

FASTAFILE2:112

gi 28279738 gb BC045535.1	100.00%	YDR120C	100.00%
gi 28279738 gb BC045535.1	100.00%		
gi 28279738 gb BC045535.1	100.00%		
gi 28279738 gb BC045535.1	100.00%		
gi 28279738 gb BC045535.1	100.00%		
Identical to NP 112196 (t-RNA methyltransferase)		t-RNA methyltransferase	

Group of orthologs #23. Best score 109 bits

Score difference with first non-orthologous sequence - FASTAFILE1:109

FASTAFILE2:109

gi 29789121 ref NM_021117	100.00%	YOR386W	100.00%
---------------------------	---------	---------	---------

Cryptochrome 2 (photolyase like) (CRY2)

Photolyase

Group of orthologs #24. Best score 105 bits

Score difference with first non-orthologous sequence - FASTAFILE1:105

FASTAFILE2:105

gi 34194217 gb BC034954.2	100.00%	YKR048C	100.00%
---------------------------	---------	---------	---------

gi 34194217 gb BC034954.2	100.00%
---------------------------	---------

Nucleosome assembly protein 1-like 3

Nucleosome assembly protein

Group of orthologs #25. Best score 80 bits

Score difference with first non-orthologous sequence - FASTAFILE1:80

FASTAFILE2:80

gi 20070937 gb BC026317.1	100.00%	YNL125C	100.00%
---------------------------	---------	---------	---------

gi 20070937 gb BC026317.1	100.00%
---------------------------	---------

**Solute carrier family 16
(monocarboxylic acid transporter)**

**Protein with similarity to
monocarboxylate permease**

Group of orthologs #26. Best score 80 bits

Score difference with first non-orthologous sequence - FASTAFILE1:80

FASTAFILE2:10

gi 45827772 ref NM_004310	100.00%	YLR262C	100.00%
---------------------------	---------	---------	---------

Ras homology gene family member H (RHOH)

**Ras-like yeast protein 2 (GTP-
binding protein)**

Group of orthologs #27. Best score 73 bits

Score difference with first non-orthologous sequence - FASTAFILE1:73

FASTAFILE2:73

gi 13111990 gb BC003168.1	100.00%	YHR073W	100.00%
---------------------------	---------	---------	---------

Oxysterol binding protein like 10

Oxysterol binding proteins homology

Group of orthologs #28. Best score 71 bits

Score difference with first non-orthologous sequence - FASTAFILE1:71

FASTAFILE2:71

gi|12652598|gb|BC000043.1 100.00% YMR138W 100.00%

ADP-ribosylation factor 4-like

GTP-binding proteins

Group of orthologs #29. Best score 68 bits

Score difference with first non-orthologous sequence - FASTAFILE1:68

FASTAFILE2:68

gi|33869323|gb|BC018823.2 100.00% YDR432W 100.00%

gi|33869323|gb|BC018823.2 100.00%

gi|33869323|gb|BC018823.2 100.00%

Splicing factor, Arg/Ser-rich 5

RNA-binding protein

Group of orthologs #30. Best score 68 bits

Score difference with first non-orthologous sequence - FASTAFILE1:68

FASTAFILE2:68

gi|50897283|ref|NM_001002 100.00% YHR189W 100.00%

gi|50897283|ref|NM_001002 100.00%

gi|50897283|ref|NM_001002 100.00%

Ribosomal protein (q9 ORF 115)

Peptidyl-Trna hydrolase

Group of orthologs #31. Best score 68 bits

Score difference with first non-orthologous sequence - FASTAFILE1:68

FASTAFILE2:68

gi|14714776|gb|BC010537.1 100.00% YMR039C 100.00%

Activated RNA polymerase II
transcription cofactor 4 (PC4)

Transcription cofactor

gi|19923783|ref|NM_006713 100.00%

PC4

Group of orthologs #32. Best score 66 bits

Score difference with first non-orthologous sequence - FASTAFILE1:66

FASTAFILE2:66

gi 38678533 ref NM_152373	100.00%	YGL254W	100.00%
---------------------------	---------	---------	---------

Zinc finger protein 684 (ZNF 684)

Transcription factor

gi 56699472 ref NM_006298	100.00%
---------------------------	---------

gi 56699472 ref NM_006298	100.00%
---------------------------	---------

gi 56699472 ref NM_006298	100.00%
---------------------------	---------

ZNF 192

gi 34303952 ref NM_152736	45.00%
---------------------------	--------

gi 34303952 ref NM_152736	45.00%
---------------------------	--------

gi 34303952 ref NM_152736	45.00%
---------------------------	--------

ZNF 187

gi 13376063 ref NM_024741	27.89%
---------------------------	--------

ZNF 408

Group of orthologs #33. Best score 66 bits

Score difference with first non-orthologous sequence - FASTAFILE1:66

FASTAFILE2:3

gi 38016126 ref NM_015014	100.00%	YNL175C	100.00%
---------------------------	---------	---------	---------

gi 38016126 ref NM_015014	100.00%
---------------------------	---------

**RNA-binding motif protein 34
(belong to RBM34 family)**

**Unknown, Nucleolar protein,
similar to NOP12P**

Group of orthologs #34. Best score 64 bits

Score difference with first non-orthologous sequence - FASTAFILE1:12

FASTAFILE2:64

gi 34783505 gb BC043346.2	100.00%	YKR094C	100.00%
---------------------------	---------	---------	---------

Ubiquitin-like 4

Ribosomal protein L40B

Group of orthologs #35. Best score 58 bits

Score difference with first non-orthologous sequence - FASTAFILE1:58

FASTAFILE2:58

gi 28416943 ref NM_003146	100.00%	YDL002C	100.00%
---------------------------	---------	---------	---------

<u>Structure specific recognition protein (SSRP1)</u>	<u>Non-histon protein (involved in chromatin remodeling)</u>
---	--

Group of orthologs #36. Best score 56 bits

**Score difference with first non-orthologous sequence - FASTAFILE1:56
FASTAFILE2:56**

gi 40789263 ref NM_032350	100.00%	YCR016W	100.00%
gi 40789263 ref NM_032350	100.00%		
gi 40789263 ref NM_032350	100.00%		
gi 40789263 ref NM_032350	100.00%		

<u>Hypothetical protein MGC11257</u>	<u>Unknown, Nuclear protein</u>
--------------------------------------	---------------------------------

Group of orthologs #37. Best score 55 bits

**Score difference with first non-orthologous sequence - FASTAFILE1:5
FASTAFILE2:55**

gi 16198524 gb BC015944.1	100.00%	YIR001C	100.00%
---------------------------	---------	---------	---------

<u>TIA1 cytotoxic granule associated RNA binding protein similar to P52912</u>	<u>RNA binding protein</u>
--	----------------------------

gi 21361701 ref NM_018107	5.56%
---------------------------	-------

RNA binding region (RNPl, RRM) containing 4

ST 5

The complete domain list of PABPs in yeast and human. The PABPs in bold demonstrate that there are more human proteins in each signaling and cellular communications domains than in yeast.

Domain_id	Number of domains in human PABPs	Number of human PABPs	Number of domains in yeast PABPs	Number of yeast PABPs	Domain name	Function group
AAA	1	1	6	6	ATPase family associated with various cellular activities	chaperone-like functions that assist in the assembly, operation, or disassembly of protein complexes
AAA_2	1	1	1	1	ATPase	perform chaperone-like functions that assist in the assembly, operation, or disassembly of protein complexes
Acetyltransf_1	1	1	1	1	Transferase	N-acetyltransferase functions
Aldo_ket_red	1	1	1	1	Aldo/keto reductase family	NADPH-dependent oxidoreductases
Aminotran_1_2	1	1	3	3	Aminotransferase class I and II	pyridoxal-phosphate dependent (PLP) amino transferase
Arf	1	1	1	1	ADP-ribosylation factor family	major regulators of vesicle biogenesis in intracellular traffic
ArfGap	1	1	2	2	Putative GTPase activating protein for Arf	Putative zinc fingers with GTPase activating proteins (GAPs) towards the small GTPase, Arf. The GAP of ARD1 stimulates GTPase hydrolysis for ARD1

						but not ARFs
Arm	4	1	8	1	Armadillo/beta-catenin-like repeat	Arm-repeat proteins function in various processes, including intracellular signaling and cytoskeletal regulation
Biotin_carb_C	2	2	1	1	Biotin carboxylase C-terminal domain	Fatty acid biosynthesis
Biotin_lipoyl	2	2	1	1	Biotin-requiring enzyme	Acyltransferase
BRCT	1	1	1	1	BRCA1 C Terminus (BRCT) domain	DNA binding
Brix	1	1	3	3	Brix domain	ribosome biogenesis and rRNA binding
C2	6	4	1	1	C2 domain	involved in signal transduction or membrane trafficking
Choline_kinase	1	1	1	1	Choline/ethanol amine kinase	catalyzes the committed step in the synthesis of phosphatidylcholine by the CDP-choline pathway Bottom of Form
CPSase_L_chain	2	2	1	1	Carbamoyl-phosphate synthase L chain, N-terminal domain	this ligase initiates both the urea cycle and the biosynthesis of arginine and/or pyrimidines
CPSase_L_D2	2	2	1	1	Carbamoyl-phosphate synthase L chain, ATP binding domain	this ligase initiates both the urea cycle and the biosynthesis of arginine and/or pyrimidines
DAO	1	1	2	2	FAD dependent oxidoreductase	includes various FAD dependent

						oxidoreductases
DBP10CT	1	1	1	1	DBP10CT (NUC160) domain	the C terminal domain is found in the Dbp10p subfamily of hypothetical RNA helicases, RNA and ATP binding, hydrolase activity
DEAD	8	8	8	8	DEAD/DEAH box helicase	RNA metabolism
DNA_photolyase	1	1	1	1	DNA photolyase	DNA repair mechanism
DnaJ	2	2	5	5	DnaJ domain	chaperone (protein folding) system
DnaJ_C	1	1	3	3	DnaJ C terminal region	Molecular chaperones
Exonuc_X-T	1	1	2	2	Exonuclease	hydrolysis of single nucleotides from the end of a DNA or RNA chain
FAD_binding_7	1	1	1	1	FAD binding domain of DNA photolyase	DNA repair enzyme
FYVE	3	3	1	1	FYVE zinc finger	DNA binding
GTP_EFTU	1	1	4	4	Elongation factor Tu GTP binding domain	catalyse the translocation of the synthesised protein chain from the A to the P site
HECT	1	1	1	1	ubiquitin-transferase	Ubiquitin-protein ligase
Helicase_C	8	8	9	9	Helicase conserved C-terminal domain	couples an ATPase activity to RNA binding and unwinding
HLH	3	3	2	2	Helix-loop-helix DNA-binding domain	DNA binding
HMG_box	2	2	1	1	HMG (high mobility group) box	bind single-stranded DNA preferentially and unwind double-

						stranded DNA
KH_1	7	5	10	2	KH domain	nucleic acid-binding
KH_2	2	1	1	1	KH domain	nucleic acid-binding
Kinesin	1	1	1	1	Kinesin motor domain	microtubule-associated force-producing protein that may play a role in organelle transport
Ldh_1_C	1	1	1	1	lactate/malate dehydrogenase, alpha/beta C-terminal domain	metabolic enzymes which catalyse the conversion of L-lactate to pyruvate
Ldh_1_N	1	1	1	1	lactate/malate dehydrogenase, NAD binding domain	metabolic enzymes which catalyse the conversion of L-lactate to pyruvate
MFS_1	1	1	2	2	Major Facilitator Superfamily	membrane transporters
MMR_HSR1	6	6	1	1	GTPase of unknown function	GTP-binding
NAP	1	1	1	1	Nucleosome assembly protein	nucleosome assembly, though to act as histone chaperones
OTU	1	1	1	1	OTU-like cysteine protease	unknown biochemical function but suggested to have cysteine protease activity
Oxysterol_BP	1	1	1	1	Oxysterol-binding protein	seem to be involved with sterol synthesis and/or its regulation
PC4	2	2	1	1	Transcriptional Coactivator p15 (PC4)	DNA binding
Pept_tRNA_hydro	1	1	1	1	Peptidyl-tRNA hydrolase	protein synthesis
PH	7	6	3	3	PH domain	intracellular signaling
PHD	2	2	2	2	PHD-finger	chromatin-mediated transcriptional

						regulation.
Pkinase	50	48	7	7	Protein kinase domain	Kinase
PseudoU_synth_1	2	1	2	1	tRNA pseudouridine synthase	tRNA processing
PseudoU_synth_2	1	1	1	1	RNA pseudouridylate synthase	RNA processing
Pyr_redox_2	1	1	1	1	Pyridine nucleotide-disulphide oxidoreductase	both class I and class II oxidoreductases and also NADH oxidases and peroxidases
Ras	2	2	3	3	Ras family	have been implicated in the regulation of hematopoietic cells, with roles in growth, survival, differentiation, cytokine production, chemotaxis, vesicle-trafficking, and phagocytosis
Rhodanese	1	1	3	3	Rhodanese-like domain	a sulphurtransferase involved in cyanide detoxification
RhoGAP	2	2	1	1	RhoGAP domain	Members of the Rho family of small G proteins transduce signals from plasma-membrane receptors and control cell adhesion, motility and shape by actin cytoskeleton formation
Ribosomal_L23	1	1	1	1	Ribosomal protein L23	protein biosynthesis
Ribosomal_L6e	1	1	1	1	Ribosomal protein L6e	protein biosynthesis
Ribosomal_L7Ae	2	2	1	1	Ribosomal protein L7Ae/L30e/S12e/Gadd45	protein biosynthesis

					family	
Ribosomal_S10	1	1	1	1	Ribosomal protein S10p/S20e	protein biosynthesis
RRM_1	23	15	10	6	RNA recognition motif	RNA binding
Rtt106	1	1	1	1	Histone chaperone Rtt106-like	a histone chaperone involved in heterochromatin-mediated silencing
S1	1	1	3	1	S1 RNA binding domain	RNA binding
Sec2p	1	1	1	1	GDP/GTP exchange factor Sec2p	required for vesicular transport at the post-Golgi stage of yeast secretion
Septin	3	3	1	1	Septin	cell cycle
SET	1	1	2	2	SET domain	contribute to epigenetic mechanisms of gene regulation by methylation of lysine residues in histones and other proteins.
SH2	3	3	1	1	SH2 domain	intracellular signaling cascade
SH3_1	9	9	3	1	SH3 domain	signal transduction
TPR_1	5	1	10	4	Tetratricopeptide repeat	mediates protein-protein interactions and the assembly of multiprotein complexes
TPR_2	5	1	1	1	Tetratricopeptide repeat	mediates protein-protein interactions and the assembly of multiprotein complexes
TRM	1	1	1	1	N2,N2-dimethylguanine tRNA methyltransferase	tRNA processing

tRNA_anti	1	1	3	3	oligonucleotide/ oligosaccharide -binding fold	nucleic acid binding
UBA	2	2	1	1	UBA/TS-N domain	are found in diverse proteins involved in the ubiquitin/proteasome pathway, DNA excision-repair, and cell signaling via protein kinases
ubiquitin	4	3	1	1	Ubiquitin family	are thought to carry a distinctive five- residue motif termed the proteasome- interacting motif (PIM), which may have a biologically significant role in protein delivery to proteasomes and recruitment of proteasomes to transcription sites
UNC-50	1	1	1	1	UNC-50 family	eukaryotic transmembrane proteins related to the Caenorhabditis elegans protein UNC- 50 . UNCL (a mammalian homologue) is a novel inner nuclear membrane protein that associates with RNA and is involved in the cell-surface expression of neuronal nicotinic receptors. UNCL plays a broader role because UNCL homologues are present in yeast and a plant species, none of which express

						nicotinic receptors and it is also found in tissues that lack nicotinic receptors.
WD40	22	7	39	8	WD domain, G-beta repeat	protein interactions
zf-C2H2	39	6	7	3	Zinc finger, C2H2 type	nucleic acid binding
zf-C3HC4	4	4	1	1	Zinc finger, C3HC4 type (RING finger)	protein-protein interactions
zf-CCCH	2	2	1	1	Zinc finger C-x8-C-x5-C-x3-H type (and similar)	DNA binding, involved in cell cycle or growth phase-related regulation
zf-CCHC	1	1	1	1	Zinc knuckle	nucleic acid binding

References

Salamat-Miller, N., Fang, J., Seidel, C.W., Smalter, A.M., and Middaugh, C.R. (September 18, 2006), A network-based analysis of polyanion binding proteins utilizing yeast protein arrays. *Mol. Cell. Proteomics* 10.1074/mcp.M600240-MCP200

Jaques, L. B. (1979). "Heparins-anionic polyelectrolyte drugs." *Pharmacol. Rev.* 31(2): 99-166.

Jaques, L. B., Bruce-Mitford, M. and Ricker, A. G. (1947). "The metachromatic activity of heparin." *Rev. Can. de Biol.* 6: 740-754.

Templeton, D., M. (1988). "General occurrence of isobestic points in the metachromatic dye complexes of sulfated glycosaminoglycans." *Int. J. Biol. Macromol.* 10: 131-136.

Wollin, A. and Jaques, L. B. (1973). "Metachromasia: an explanation of the color change produced in dyes by heparin and other substances." *Thromb Res* 2: 377-382.